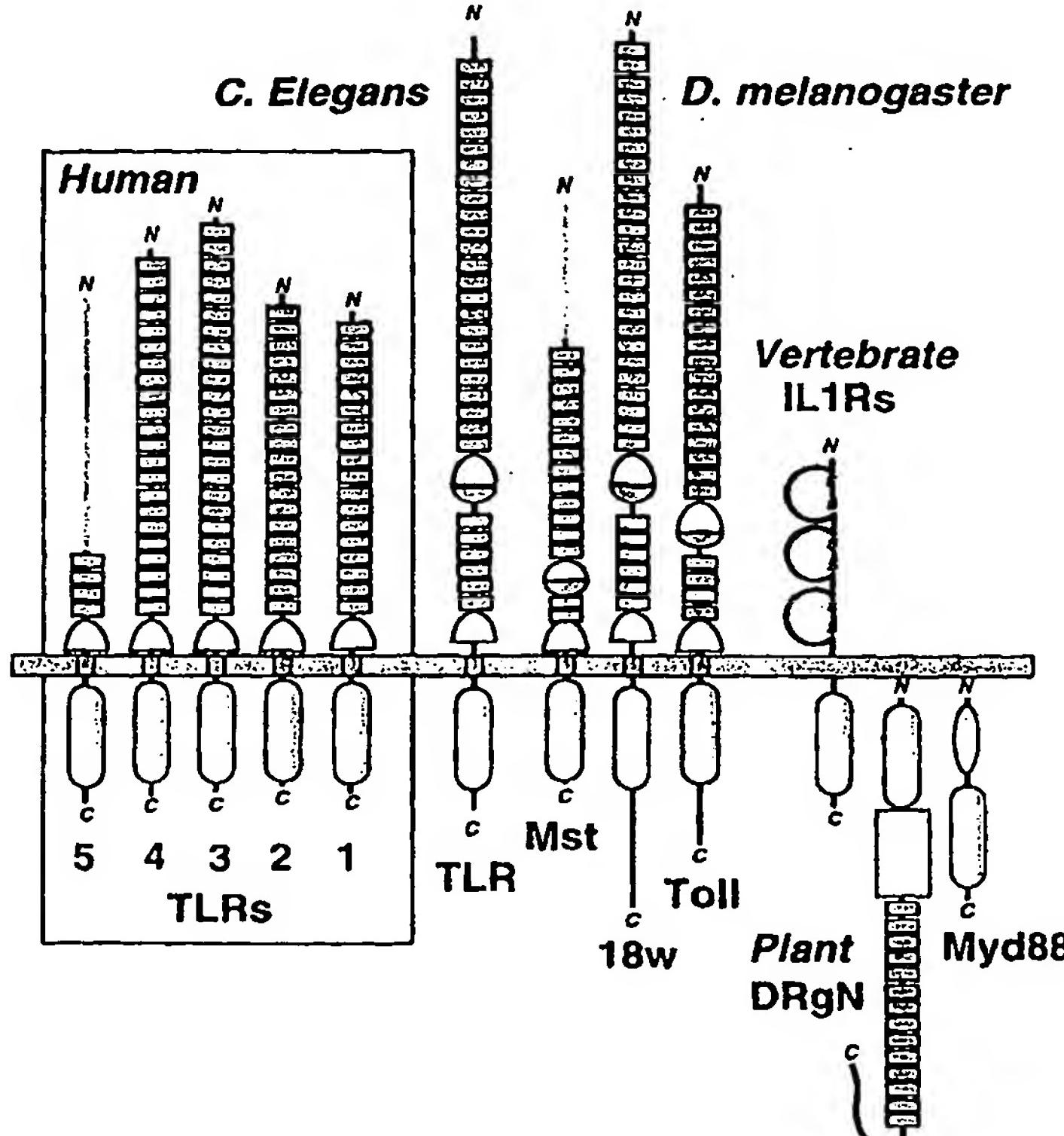




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :  C12N 15/12, 15/62, C07K 14/705, 16/28, C12N 15/70		A2	(11) International Publication Number: <b>WO 98/50547</b>  (43) International Publication Date: 12 November 1998 (12.11.98)									
<p>(21) International Application Number: PCT/US98/08979</p> <p>(22) International Filing Date: 7 May 1998 (07.05.98)</p> <p>(30) Priority Data:</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">60/044,293</td> <td style="width: 25%;">7 May 1997 (07.05.97)</td> <td style="width: 25%;">US</td> </tr> <tr> <td>60/072,212</td> <td>22 January 1998 (22.01.98)</td> <td>US</td> </tr> <tr> <td>60/076,947</td> <td>5 March 1998 (05.03.98)</td> <td>US</td> </tr> </table> <p>(71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033 (US).</p> <p>(72) Inventors: HARDIMAN, Gerard, T.; 4 Howe Street, Watertown, MA 02172 (US). ROCK, Fernando, L.; 721 Shell Boulevard #203, Foster City, CA 94404 (US). BAZAN, J., Fernando; 775 University Drive, Menlo Park, CA 94025 (US). KASTELEIN, Robert, A.; 463 Summit Drive, Redwood City, CA 94062 (US).</p> <p>(74) Agents: McLAUGHLIN, Jaye, P. et al.; Schering-Plough Corporation, Patent Dept. K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).</p>		60/044,293	7 May 1997 (07.05.97)	US	60/072,212	22 January 1998 (22.01.98)	US	60/076,947	5 March 1998 (05.03.98)	US	<p>(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CZ, EE, GE, GW, HU, ID, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i></p>	
60/044,293	7 May 1997 (07.05.97)	US										
60/072,212	22 January 1998 (22.01.98)	US										
60/076,947	5 March 1998 (05.03.98)	US										
<p>(54) Title: HUMAN TOLL-LIKE RECEPTOR PROTEINS, RELATED REAGENTS AND METHODS</p> <p>(57) Abstract</p> <p>Nucleic acids encoding nine human receptors, designated DNAX Toll-like receptors 2–10 (DTLR2–10), homologous to the Drosophila Toll receptor and the human IL-1 receptor, purified DTLR proteins and fragments thereof, monoclonal antibodies against these receptors, and methods for diagnostic and therapeutic use.</p>												



***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

HUMAN RECEPTOR PROTEINS; RELATED REAGENTS AND METHODS

This filing claims priority from U.S. Patent  
5 Applications USSN 60/044,293, filed May 7, 1997; USSN  
60/072,212, filed January 22, 1998; and USSN 60/076,947,  
filed March 5, 1998, each of which is incorporated herein  
by reference.

10

FIELD OF THE INVENTION

The present invention relates to compositions and  
methods for affecting mammalian physiology, including  
morphogenesis or immune system function. In particular,  
it provides nucleic acids, proteins, and antibodies which  
15 regulate development and/or the immune system.  
Diagnostic and therapeutic uses of these materials are  
also disclosed.

20

BACKGROUND OF THE INVENTION

Recombinant DNA technology refers generally to  
techniques of integrating genetic information from a  
donor source into vectors for subsequent processing, such  
as through introduction into a host, whereby the  
transferred genetic information is copied and/or  
25 expressed in the new environment. Commonly, the genetic  
information exists in the form of complementary DNA  
(cDNA) derived from messenger RNA (mRNA) coding for a  
desired protein product. The carrier is frequently a  
plasmid having the capacity to incorporate cDNA for later  
30 replication in a host and, in some cases, actually to  
control expression of the cDNA and thereby direct  
synthesis of the encoded product in the host.

For some time, it has been known that the mammalian  
immune response is based on a series of complex cellular  
35 interactions, called the "immune network". Recent  
research has provided new insights into the inner  
workings of this network. While it remains clear that

much of the immune response does, in fact, revolve around the network-like interactions of lymphocytes, macrophages, granulocytes, and other cells, immunologists now generally hold the opinion that soluble proteins, known as lymphokines, cytokines, or monokines, play critical roles in controlling these cellular interactions. Thus, there is considerable interest in the isolation, characterization, and mechanisms of action of cell modulatory factors, an understanding of which will lead to significant advancements in the diagnosis and therapy of numerous medical abnormalities, e.g., immune system disorders.

Lymphokines apparently mediate cellular activities in a variety of ways. They have been shown to support the proliferation, growth, and/or differentiation of pluripotential hematopoietic stem cells into vast numbers of progenitors comprising diverse cellular lineages which make up a complex immune system. Proper and balanced interactions between the cellular components are necessary for a healthy immune response. The different cellular lineages often respond in a different manner when lymphokines are administered in conjunction with other agents.

Cell lineages especially important to the immune response include two classes of lymphocytes: B-cells, which can produce and secrete immunoglobulins (proteins with the capability of recognizing and binding to foreign matter to effect its removal), and T-cells of various subsets that secrete lymphokines and induce or suppress the B-cells and various other cells (including other T-cells) making up the immune network. These lymphocytes interact with many other cell types.

Another important cell lineage is the mast cell (which has not been positively identified in all mammalian species), which is a granule-containing connective tissue cell located proximal to capillaries throughout the body. These cells are found in especially

high concentrations in the lungs, skin, and gastrointestinal and genitourinary tracts. Mast cells play a central role in allergy-related disorders, particularly anaphylaxis as follows: when selected 5 antigens crosslink one class of immunoglobulins bound to receptors on the mast cell surface, the mast cell degranulates and releases mediators, e.g., histamine, serotonin, heparin, and prostaglandins, which cause allergic reactions, e.g., anaphylaxis.

10 Research to better understand and treat various immune disorders has been hampered by the general inability to maintain cells of the immune system in vitro. Immunologists have discovered that culturing many of these cells can be accomplished through the use of T-cell and other cell supernatants, which contain various growth factors, including many of the lymphokines.

The interleukin-1 family of proteins includes the IL-1 $\alpha$ , the IL-1 $\beta$ , the IL-1RA, and recently the IL-1 $\gamma$  (also designated Interferon-Gamma Inducing Factor, IGIF). 20 This related family of genes have been implicated in a broad range of biological functions. See Dinarello (1994) FASEB J. 8:1314-1325; Dinarello (1991) Blood 77:1627-1652; and Okamura, et al. (1995) Nature 378:88-91.

25 In addition, various growth and regulatory factors exist which modulate morphogenetic development. This includes, e.g., the Toll ligands, which signal through binding to receptors which share structural, and mechanistic, features characteristic of the IL-1 30 receptors. See, e.g., Lemaitre, et al. (1996) Cell 86:973-983; and Belvin and Anderson (1996) Ann. Rev. Cell & Devel. Biol. 12:393-416.

From the foregoing, it is evident that the discovery and development of new soluble proteins and their 35 receptors, including ones similar to lymphokines, should contribute to new therapies for a wide range of degenerative or abnormal conditions which directly or

indirectly involve development, differentiation, or function, e.g., of the immune system and/or hematopoietic cells. In particular, the discovery and understanding of novel receptors for lymphokine-like molecules which  
5 enhance or potentiate the beneficial activities of other lymphokines would be highly advantageous. The present invention provides new receptors for ligands exhibiting similarity to interleukin-1 like compositions and related compounds, and methods for their use.

10

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic comparison of the protein architectures of Drosophila and human DTLRs, and their relationship to vertebrate IL-1 receptors and plant  
15 disease resistance proteins. Three Drosophila (Dm) DTLRs (Toll, 18w, and the Mst ORF fragment) (Morisato and Anderson (1995) *Ann. Rev. Genet.* 29:371-399; Chiang and Beachy (1994) *Mech. Develop.* 47:225-239; Mitcham, et al.  
1996) *J. Biol. Chem.* 271:5777-5783; and Eldon, et al.  
20 (1994) *Develop.* 120:885-899) are arrayed beside four complete (DTLRs 1-4) and one partial (DTLR5) human (Hu) receptors. Individual LRRs in the receptor ectodomains that are flagged by PRINTS (Attwood, et al. (1997) *Nucleic Acids Res.* 25:212-217) are explicitly noted by  
25 boxes; 'top' and 'bottom' Cys-rich clusters that flank the C- or N-terminal ends of LRR arrays are respectively drawn by apposed half-circles. The loss of the internal Cys-rich region in DTLRs 1-5 largely accounts for their smaller ectodomains (558, 570, 690, and 652 aa,  
30 respectively) when compared to the 784 and 977 aa extensions of Toll and 18w. The incomplete chains of DmMst and HuDTLR5 (519 and 153 aa ectodomains, respectively) are represented by dashed lines. The intracellular signaling module common to DTLRs, IL-1-type  
35 receptors (IL-1Rs), the intracellular protein Myd88, and the tobacco disease resistance gene N product (DRgN) is indicated below the membrane. See, e.g., Hardiman, et

al. (1996) Oncogene 13:2467-2475; and Rock, et al. (1998) Proc. Nat'l Acad. Sci. USA 95:588-. Additional domains include the trio of Ig-like modules in IL-1Rs (disulfide-linked loops); the DRgN protein features an NTPase domain 5 (box) and Myd88 has a death domain (black oval).

Figures 2A-2B show conserved structural patterns in the signaling domains of Toll- and IL-1-like cytokine receptors, and two divergent modular proteins. Figure 2A shows a sequence alignment of the common TH domain.

10 DTLRs are labeled as in Figure 1; the human (Hu) or mouse (Mo) IL-1 family receptors (IL-1R1-6) are sequentially numbered as earlier proposed (Hardiman, et al. (1996) Oncogene 13:2467-2475); Myd88 and the sequences from tobacco (To) and flax, *L. usitatissimum* (Lu), represent

15 C- and N-terminal domains, respectively, of larger, multidomain molecules. Ungapped blocks of sequence (numbered 1-10) are boxed. Triangles indicate deleterious mutations, while truncations N-terminal of the arrow eliminate bioactivity in human IL-1R1 (Heguy, et al. (1992) J. Biol. Chem. 267:2605-2609). PHD (Rost and Sander (1994) Proteins 19:55-72) and DSC (King and Sternberg (1996) Protein Sci. 5:2298-2310) secondary structure predictions of  $\alpha$ -helix (H),  $\beta$ -strand (E), or coil (L) are marked. The amino acid shading scheme

20 depicts chemically similar residues: hydrophobic, acidic, basic, Cys, aromatic, structure-breaking, and tiny.

25 Diagnostic sequence patterns for IL-1Rs, DTLRs, and full alignment (ALL) were derived by Consensus at a stringency of 75%. Symbols for amino acid subsets are (see internet site for detail): o, alcohol; l, aliphatic; ., any amino acid; a, aromatic; c, charged; h, hydrophobic; -, negative; p, polar; +, positive; s, small; u, tiny; t, turnlike. Figure 2B shows a topology diagram of the proposed TH  $\beta/\alpha$  domain fold. The parallel  $\beta$ -sheet (with

30 35  $\beta$ -strands A-E as yellow triangles) is seen at its C-terminal end;  $\alpha$ -helices (circles labeled 1-5) link the  $\beta$ -strands; chain connections are to the front (visible) or

back (hidden). Conserved, charged residues at the C-end of the  $\beta$ -sheet are noted in gray (Asp) or as a lone black (Arg) residue (see text).

Figure 3 shows evolution of a signaling domain superfamily. The multiple TH module alignment of Figure 2A was used to derive a phylogenetic tree by the Neighbor-Joining method (Thompson, et al. (1994) Nucleic Acids Res. 22:4673-4680). Proteins labeled as in the alignment; the tree was rendered with TreeView.

Figures 4A-4D show FISH chromosomal mapping of human DTLR genes. Denatured chromosomes from synchronous cultures of human lymphocytes were hybridized to biotinylated DTLR cDNA probes for localization. The assignment of the FISH mapping data (left, Figures 4A, 10 DTLR2; 4B, DTLR3; 4C, DTLR4; 4D, DTLR5) with chromosomal bands was achieved by superimposing FISH signals with 15 DAPI banded chromosomes (center panels). Heng and Tsui (1994) Meth. Molec. Biol. 33:109-122. Analyses are summarized in the form of human chromosome ideograms 20 (right panels).

Figures 5A-5F show mRNA blot analyses of Human DTLRs. Human multiple tissue blots (He, heart; Br, brain; Pl, placenta; Lu, lung; Li, liver; Mu, muscle; Ki, kidney; Pn, Pancreas; Sp, spleen; Th, thymus; Pr, 25 prostate; Te, testis; Ov, ovary, SI, small intestine; Co, colon; PBL, peripheral blood lymphocytes) and cancer cell line (promyelocytic leukemia, HL60; cervical cancer, HELAS3; chronic myelogenous leukemia, K562; lymphoblastic leukemia, Molt4; colorectal adenocarcinoma, SW480; 30 melanoma, G361; Burkitt's Lymphoma Raji, Burkitt's; colorectal adenocarcinoma, SW480; lung carcinoma, A549) containing approximately 2  $\mu$ g of poly(A)<sup>+</sup> RNA per lane were probed with radiolabeled cDNAs encoding DTLR1 (Figures 5A-5C), DTLR2 (Figure 5D), DTLR3 (Figure 5E), 35 and DTLR4 (Figure 5F) as described. Blots were exposed to X-ray film for 2 days (Figures 5A-5C) or one week (Figure 5D-5F) at -70° C with intensifying screens. An

anomalous 0.3 kB species appears in some lanes; hybridization experiments exclude a message encoding a DTLR cytoplasmic fragment.

SUMMARY OF THE INVENTION

- 5       The present invention is directed to nine novel related mammalian receptors, e.g., human, Toll receptor like molecular structures, designated DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, and DTLR10, and their biological activities. It includes nucleic acids  
10 coding for the polypeptides themselves and methods for their production and use. The nucleic acids of the invention are characterized, in part, by their homology to cloned complementary DNA (cDNA) sequences enclosed herein.
- 15      In certain embodiments, the invention provides a composition of matter selected from the group of: a substantially pure or recombinant DTLR2 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 4; a natural sequence DTLR2 of SEQ ID NO: 4; a fusion protein comprising DTLR2 sequence; a substantially pure or recombinant DTLR3 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 6; a natural  
20 sequence DTLR3 of SEQ ID NO: 6; a fusion protein comprising DTLR3 sequence; a substantially pure or recombinant DTLR4 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 26; a natural sequence  
25 DTLR4 of SEQ ID NO: 26; a fusion protein comprising DTLR4 sequence; a substantially pure or recombinant DTLR5 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 10; a natural sequence DTLR5 of SEQ ID NO:  
30 35 10; and a fusion protein comprising DTLR5 sequence.

In other embodiments, the invention provides a composition of matter selected from the group of: a

substantially pure or recombinant DTLR6 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 12; a natural sequence DTLR6 of SEQ ID NO: 12; a  
5 fusion protein comprising DTLR6 sequence; a substantially pure or recombinant DTLR7 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 16 or 18 or; a natural sequence DTLR7 of SEQ ID NO: 16 or 18; a fusion  
10 protein comprising DTLR7 sequence; a substantially pure or recombinant DTLR8 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 32; a natural sequence DTLR8 of SEQ ID NO: 32; a fusion protein  
15 comprising DTLR8 sequence; a substantially pure or recombinant DTLR9 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 22; a natural sequence DTLR9 of SEQ ID NO: 22; and a fusion protein comprising  
20 DTLR9 sequence; a substantially pure or recombinant DTLR10 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 34; a natural sequence DTLR10 of SEQ ID NO: 34; and a fusion protein comprising DTLR10  
25 sequence.

Preferably, the substantially pure or isolated protein comprises a segment exhibiting sequence identity to a corresponding portion of a DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10, wherein:  
30 the homology is at least about 90% identity and the portion is at least about 9 amino acids; the homology is at least about 80% identity and the portion is at least about 17 amino acids; or the homology is at least about 70% identity and the portion is at least about 25 amino  
35 acids. In specific embodiments, the composition of matter: is DTLR2, which comprises a mature sequence of SEQ ID NO: 4; or exhibits a post-translational+

modification pattern distinct from natural DTLR2; is  
DTLR3, which comprises a mature sequence of SEQ ID NO: 6;  
or exhibits a post-translational modification pattern  
distinct from natural DTLR3; is DTLR4, which: comprises a  
5 mature sequence of SEQ ID NO: 26; or exhibits a post-  
translational modification pattern distinct from natural  
DTLR4; or is DTLR5, which: comprises the complete  
sequence of SEQ ID NO: 10; or exhibits a post-  
translational modification pattern distinct from natural  
10 DTLR5; or is DTLR6, which comprises a mature sequence of  
SEQ ID NO: 12; or exhibits a post-translational  
modification pattern distinct from natural DTLR6; is  
DTLR7, which comprises a mature sequence of SEQ ID NO: 16  
or 18; or exhibits a post-translational modification  
15 pattern distinct from natural DTLR7; is DTLR8, which:  
comprises a mature sequence of SEQ ID NO: 32; or exhibits  
a post-translational modification pattern distinct from  
natural DTLR8; or is DTLR9, which: comprises the complete  
sequence of SEQ ID NO: 22; or exhibits a post-  
20 translational modification pattern distinct from natural  
DTLR9; or is DTLR10, which: comprises the complete  
sequence of SEQ ID NO: 34; or exhibits a post-  
translational modification pattern distinct from natural  
DTLR10; or the composition of matter may be a protein or  
25 peptide which: is from a warm blooded animal selected  
from a mammal, including a primate, such as a human;  
comprises at least one polypeptide segment of SEQ ID NO:  
4, 6, 26, 10, 12, 16, 18, 32, 22 or 34; exhibits a  
plurality of portions exhibiting said identity; is a  
30 natural allelic variant of DTLR2, DTLR3, DTLR4, DTLR5,  
DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10; has a length at  
least about 30 amino acids; exhibits at least two non-  
overlapping epitopes which are specific for a primate  
DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9,  
35 or DTLR10; exhibits a sequence identity at least about  
90% over a length of at least about 20 amino acids to a  
primate DTLR2, DTLR3, DTLR4, DTLR5, DTLR6; exhibits at

least two non-overlapping epitopes which are specific for a primate DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10; exhibits a sequence identity at least about 90% over a length of at least about 20 amino acids to a primate DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10; is glycosylated; has a molecular weight of at least 100 kD with natural glycosylation; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence.

Other embodiments include a composition comprising: a sterile DTLR2 protein or peptide; or the DTLR2 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR3 protein or peptide; or the DTLR3 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR4 protein or peptide; or the DTLR4 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR5 protein or peptide; or the DTLR5 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR6 protein or peptide; or the DTLR6 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR7 protein or peptide; or the DTLR7 protein or peptide and a carrier,

wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR8 protein or peptide; or the DTLR8 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR9 protein or peptide; or the DTLR9 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile DTLR10 protein or peptide; or the DTLR10 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

In certain fusion protein embodiments, the invention provides a fusion protein comprising: mature protein sequence of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34; a detection or purification tag, including a FLAG, His6, or Ig sequence; or sequence of another receptor protein.

Various kit embodiments include a kit comprising a DTLR protein or polypeptide, and: a compartment comprising the protein or polypeptide; and/or instructions for use or disposal of reagents in the kit.

Binding compound embodiments include those comprising an antigen binding site from an antibody, which specifically binds to a natural DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10 protein, wherein: the protein is a primate protein; the binding compound is an Fv, Fab, or Fab2 fragment; the binding compound is conjugated to another chemical moiety; or the antibody: is raised against a peptide sequence of a mature polypeptide of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34; is raised against a mature

DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9 or DTLR10; is raised to a purified human DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9 or DTLR10; is immunoselected; is a polyclonal antibody; binds to a  
5 denatured DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9 or DTLR10; exhibits a Kd to antigen of at least 30  $\mu$ M; is attached to a solid substrate, including a bead or plastic membrane; is in a sterile composition; or is detectably labeled, including a radioactive or  
10 fluorescent label. A binding composition kit often comprises the binding compound, and: a compartment comprising said binding compound; and/or instructions for use or disposal of reagents in the kit. Often the kit is capable of making a qualitative or quantitative analysis.

15 Other compositions include a composition comprising: a sterile binding compound, or the binding compound and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral  
20 administration.

Nucleic acid embodiments include an isolated or recombinant nucleic acid encoding a DTLR2-10 protein or peptide or fusion protein, wherein: the DTLR is from a mammal; or the nucleic acid: encodes an antigenic peptide  
25 sequence of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34; exhibits at least about 80% identity to a natural cDNA encoding said segment; is an expression  
30 vector; further comprises an origin of replication; is from a natural source; comprises a detectable label; comprises synthetic nucleotide sequence; is less than 6 kb, preferably less than 3 kb; is from a mammal, including a primate; comprises a natural full length  
35 coding sequence; is a hybridization probe for a gene encoding said DTLR; or is a PCR primer, PCR product, or mutagenesis primer. A cell, tissue, or organ comprising

such a recombinant nucleic acid is also provided. Preferably, the cell is: a prokaryotic cell; a eukaryotic cell; a bacterial cell; a yeast cell; an insect cell; a mammalian cell; a mouse cell; a primate cell; or a human cell. Kits are provided comprising such nucleic acids, and: a compartment comprising said nucleic acid; a compartment further comprising a primate DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9 or DTLR10 protein or polypeptide; and/or instructions for use or disposal of reagents in the kit. Often, the kit is capable of making a qualitative or quantitative analysis.

Other embodiments include a nucleic acid which: hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 3; hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 5; hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 25; hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 9; hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 11; hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 15 or 17; hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 31; hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 21; hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 33; exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a primate DTLR2 DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9 or DTLR10.

Preferably, such nucleic acid will have such properties, wherein: wash conditions are at 45° C and/or 500 mM salt; or the identity is at least 90% and/or the stretch is at least 55 nucleotides. More preferably, the wash conditions are at 55° C and/or 150 mM salt; or the identity is at least 95% and/or the stretch is at least 75 nucleotides.

The invention also provides a method of modulating physiology or development of a cell or tissue culture cells comprising contacting the cell with an agonist or antagonist of a mammalian DTLR2, DTLR3, DTLR4, DTLR5, 5 DTLR6, DTLR7, DTLR8, DTLR9, or DTLR10.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. General

10 The present invention provides the amino acid sequence and DNA sequence of mammalian, herein primate DNAX Toll like receptor molecules (DTLR) having particular defined properties, both structural and biological. These have been designated herein as DTLR2, 15 DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, and DTLR10, respectively, and increase the number of members of the human Toll like receptor family from 1 to 10. Various cDNAs encoding these molecules were obtained from primate, e.g., human, cDNA sequence libraries. Other 20 primate or other mammalian counterparts would also be desired.

Some of the standard methods applicable are described or referenced, e.g., in Maniatis, et al. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring 25 Harbor Laboratory, Cold Spring Harbor Press; Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual, (2d ed.), vols 1-3, CSH Press, NY; Ausubel, et al., Biology, Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (1987 and periodic supplements) Current Protocols 30 in Molecular Biology, Greene/Wiley, New York; each of which is incorporated herein by reference.

A complete nucleotide and corresponding amino acid sequence of a human DTLR1 coding segment is shown in SEQ ID NO: 1 and 2. See also Nomura, et al. (1994) DNA Res 35 1:27-35. A complete nucleotide and corresponding amino acid sequence of a human DTLR2 coding segment is shown in SEQ ID NO: 3 and 4. A complete nucleotide and

corresponding amino acid sequence of a human DTLR3 coding segment is shown in SEQ ID NO: 5 and 6. A complete nucleotide and corresponding amino acid sequence of a human DTLR4 coding segment is shown in SEQ ID NO: 7 and 8. An alternate nucleic acid and corresponding amino acid sequence of a human DTLR4 coding segment is provided in SEQ ID NO: 25 and 26. A partial nucleotide and corresponding amino acid sequence of a human DTLR5 coding segment is shown in SEQ ID NO: 9 and 10. A complete nucleotide and corresponding amino acid sequence of a human DTLR6 coding segment is shown in SEQ ID NO: 11 and 12 and a partial sequence of a mouse DTLR6 is provided in SEQ ID NO: 13 and 14. Additional mouse DTLR6 sequence is provided in SEQ ID NO: 27 and 29 (nucleotide sequence) and SEQ ID NO: 28 and 30 (amino acid sequence). Partial nucleotide (SEQ ID NO: 15 and 17) and corresponding amino acid sequence (SEQ ID NO: 16 and 18) of a human DTLR7 coding segment is also provided. Partial nucleotide and corresponding amino acid sequence of a human DTLR8 coding segment is shown in SEQ ID NO: 19 and 20. A more complete nucleotide and corresponding amino acid sequence of a human DTLR coding segment is shown in SEQ ID NO: 31 and 32. Partial nucleotide and corresponding amino acid sequence of a human DTLR9 coding segment is shown in SEQ ID NO: 21 and 22. Partial nucleotide and corresponding amino acid sequence of a human DTLR10 coding segment is shown in SEQ ID NO: 23 and 24. More complete nucleotide and corresponding amino acid sequence of a human DTLR10 coding segment is shown in SEQ ID NO: 33 and 34. A partial nucleotide sequence for a mouse DTLR10 coding segment is provided in SEQ ID NO: 35.

Table 1: Comparison of intracellular domains of human DTLRs. DTLR1 is SEQ ID NO: 2; DTLR2 is SEQ ID NO: 4; DTLR3 is SEQ ID NO: 6; DTLR4 is SEQ ID NO: 8; DTLR5 is SEQ ID NO: 10; and DTLR6 is SEQ ID NO: 12. Particularly important and conserved, e.g., characteristic, residues correspond, across the DTLRs, to SEQ ID NO: 18 residues tyr10-tyr13; trp26; cys46; trp52; pro54-gly55; ser69; lys71; trp134-pro135; and phe144-trp145.

	DTLR1	QRNLQFHAFISYSQHD---SFWVKNELLPNLEKEG-----MQICLHERNF
10	DTLR9	KENLQFHAFISYSQHED---SAWKSELVPYLEKED-----IQICLHERNF
	DTLR8	-----NELIPNLEKEDGS---ILICLYESYF
	DTLR2	SRNICYDAFVSYSERD---AYWVENLMVQELENFNPP---FKLCLHKRDF
	DTLR6	SPDCCYDAFIVYDTKDPAVTEWLAELVAKLEDPREK--HFNLCLEERDW
	DTLR7	TSQTFYDAYISYDTKDASVTWDVINELYHLEESRDK--NVLLCLEERDW
15	DTLR10	EDALPYDAFVVFDKTXSAADVYNELRGQLEECRGRW-ALRLCLEERDW
	DTLR4	RGENIYDAFVIYSSQD---EDWVRNELVKNLEEGVPP---FQLCLHYRDF
	DTLR5	PDMYKYDAYLCFSSKD---FTWVQNALLKHLDTQYSDQNRNFNLCFEERDF
	DTLR3	TEQFEYAAIIHAYKD---KDWVWEHFSSMEKEDQS---LKFCLEERDF
		: . . . : * : :
20	DTLR1	VPGKSIVENIITC-IEKSYKSIFVLSPNFVQSEWCH-YELYFAHHNLFHE
	DTLR9	VPGKSIVENIINC-IEKSYKSIFVLSPNFVQSEWCH-YELYFAHHNLFHE
	DTLR8	DPGKSISENIVSF-IEKSYKSIFVLSPNFVQNEWCH-YEFYFAHHNLFHE
	DTLR2	IPGKWIIDNIIDS-IEKSHKTVFVLSNFVSEWCK-YELDFSHFRLFEE
25	DTLR6	LPGQPVLNLSQS-IQLSKKTVFVMTDKYAKTENFK-IAFYLSHQRLMDE
	DTLR7	DPLGLAIIDNLMQS-INQSKKTVFVLTCKYAKSWNFK-TAFYXLQRLMGE
	DTLR10	LPGKTLFENLWAS-VYGSRKTLFVLAHTDRVSGLLR-AIFLLAQQRILLE-
	DTLR4	IPGVAIAANIIHEGFHRSRKVIVVVSQHFIQSRWC-FYEIAQTWQFLS
	DTLR5	VPGENRIANIQDA-IWNSRKIVCLVSRHFLRGWCL-EAFSYAQGRCLSD
30	DTLR3	EAGVFELEAIVNS-IKRSRKIIFVITHHLLKDPLCKRFKVHHAVQQAIEQ
		. * : . * * : :: :
	DTLR1	GSNSLILILLEPIPIPQYSIPSSYHKLKSLMARRTYLEWPKEKSKRGLFWAN
	DTLR9	GSNNLILILLEPIPIPQNSIPNKYHKLKALMTQRTYLQWPKEKSKRGLFWA-
35	DTLR8	NSDHILLILEPIPFYCIPTRYHkleallekkaylewpkdrkcglnfwan
	DTLR2	NNDAAILILLEPIEKKAIPQRFCRKIMNTKTYLEWPMDEAQREGFWVN
	DTLR6	KVDVIIILIFLEKPFQK---SKFLQLRKRLCGSSVLEWPTNPQAHPYFWQC
	DTLR7	NMDVIIIFILLEPVLQH---SPYLRLRQICKSSILQWPDPNPKAERLFWQT
	DTLR10	-
40	DTLR4	SRAGIIFIVLQKVEKT-LLRQQVELYRLLSRNTYLEWEDSVLGRHIFWRR
	DTLR5	LNSALIMVVVGSLSQY-QLMKHQSIIRGVQKQQYLRWPEDLQDVGFHLHK
	DTLR3	NLDIILVFLEEIPDYKLNHALCLRRGMFKSHCILNWPVQKERIGAFRHK
45	DTLR1	LRAAINIKLTEQAKK-----
	DTLR9	-----
	DTLR8	LRAAVNVNVLATREMYELQTFTELNEESRGSTISLMRTDCL
	DTLR2	LRAAIKS-----
	DTLR6	LKNALATDNHVAYSQVFKETV-----
50	DTLR7	LXNVVLTENDSRYNNMYVDSIKQY-----
	DTLR10	-----
	DTLR4	LRKALLDGKSWNPEGTVGTGCNWQEATSI-----
	DTLR5	LSQQILKKEKEKKKDNNIPLQTVATIS-----
	DTLR3	LQVALGSKNSVH-----

As used herein, the term DNAX Toll like receptor 2 (DTLR2) shall be used to describe a protein comprising a protein or peptide segment having or sharing the amino acid sequence shown in SEQ ID NO: 4, or a substantial fragment thereof. Similarly, with a DTLR3 and SEQ ID NO: 5; DTLR4 and SEQ ID NO: 26; DTLR5 and SEQ ID NO: 10; DTLR6 and SEQ ID NO: 12; DTLR7 and SEQ ID NO: 16 and 18; DTLR8 and SEQ ID NO: 32; DTLR9 and SEQ ID NO: 22; and 10 DTLR10 and SEQ ID NO: 34.

The invention also includes a protein variations of the respective DTLR allele whose sequence is provided, e.g., a mutein agonist or antagonist. Typically, such agonists or antagonists will exhibit less than about 10% sequence differences, and thus will often have between 1- and 11-fold substitutions, e.g., 2-, 3-, 5-, 7-fold, and others. It also encompasses allelic and other variants, e.g., natural polymorphic, of the protein described. Typically, it will bind to its corresponding biological receptor with high affinity, e.g., at least about 100 nM, usually better than about 30 nM, preferably better than about 10 nM, and more preferably at better than about 3 nM. The term shall also be used herein to refer to related naturally occurring forms, e.g., alleles, 20 polymorphic variants, and metabolic variants of the mammalian protein.

This invention also encompasses proteins or peptides having substantial amino acid sequence identity with the amino acid sequence in SEQ ID NO: 4. It will include 30 sequence variants with relatively few substitutions, e.g., preferably less than about 3-5. Similar features apply to the other DTLR sequences provided in SEQ ID NO: 6, 26, 10, 12, 16, 18, 32, 22 and 34.

A substantial polypeptide "fragment", or "segment", 35 is a stretch of amino acid residues of at least about 8 amino acids, generally at least 10 amino acids, more generally at least 12 amino acids, often at least 14

amino acids, more often at least 16 amino acids, typically at least 18 amino acids, more typically at least 20 amino acids, usually at least 22 amino acids, more usually at least 24 amino acids, preferably at least 5 26 amino acids, more preferably at least 28 amino acids, and, in particularly preferred embodiments, at least about 30 or more amino acids. Sequences of segments of different proteins can be compared to one another over appropriate length stretches.

10 Amino acid sequence homology, or sequence identity, is determined by optimizing residue matches, if necessary, by introducing gaps as required. See, e.g., Needleham, et al., (1970) J. Mol. Biol. 48:443-453; Sankoff, et al., (1983) chapter one in Time Warps, String  
15 Edits, and Macromolecules: The Theory and Practice of Sequence Comparsion, Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA; and the University of Wisconsin Genetics Computer Group (GCG), Madison, WI; each of which is incorporated  
20 herein by reference. This changes when considering conservative substitutions as matches. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine,  
25 glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Homologous amino acid sequences are intended to include natural allelic and interspecies variations in the cytokine sequence. Typical homologous proteins or peptides will have from 50-100% homology (if  
30 gaps can be introduced), to 60-100% homology (if conservative substitutions are included) with an amino acid sequence segment of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34. Homology measures will be at least about 70%, generally at least 76%, more generally at  
35 least 81%, often at least 85%, more often at least 88%, typically at least 90%, more typically at least 92%, usually at least 94%, more usually at least 95%,

preferably at least 96%, and more preferably at least 97%, and in particularly preferred embodiments, at least 98% or more. The degree of homology will vary with the length of the compared segments. Homologous proteins or 5 peptides, such as the allelic variants, will share most biological activities with the embodiments described in SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34. Particularly interesting regions of comparison, at the amino acid or nucleotide levels, correspond to those 10 within each of the blocks 1-10, or intrablock regions, corresponding to those indicated in Figure 2A.

As used herein, the term "biological activity" is used to describe, without limitation, effects on inflammatory responses, innate immunity, and/or 15 morphogenic development by respective ligands. For example, these receptors should, like IL-1 receptors, mediate phosphatase or phosphorylase activities, which activities are easily measured by standard procedures. See, e.g., Hardie, et al. (eds. 1995) The Protein Kinase 20 FactBook vols. I and II, Academic Press, San Diego, CA; Hanks, et al. (1991) Meth. Enzymol. 200:38-62; Hunter, et al. (1992) Cell 70:375-388; Lewin (1990) Cell 61:743-752; Pines, et al. (1991) Cold Spring Harbor Symp. Quant. Biol. 56:449-463; and Parker, et al. (1993) Nature 25 363:736-738. The receptors exhibit biological activities much like regulatable enzymes, regulated by ligand binding. However, the enzyme turnover number is more close to an enzyme than a receptor complex. Moreover, the numbers of occupied receptors necessary to induce 30 such enzymatic activity is less than most receptor systems, and may number closer to dozens per cell, in contrast to most receptors which will trigger at numbers in the thousands per cell. The receptors, or portions thereof, may be useful as phosphate labeling enzymes to 35 label general or specific substrates.

The terms ligand, agonist, antagonist, and analog of, e.g., a DTLR, include molecules that modulate the

characteristic cellular responses to Toll ligand like proteins, as well as molecules possessing the more standard structural binding competition features of ligand-receptor interactions, e.g., where the receptor is 5 a natural receptor or an antibody. The cellular responses likely are mediated through binding of various Toll ligands to cellular receptors related to, but possibly distinct from, the type I or type II IL-1 receptors. See, e.g., Belvin and Anderson (1996) Ann. 10 Rev. Cell Dev. Biol. 12:393-416; Morisato and Anderson (1995) Ann. Rev. Genetics 29:371-3991 and Hultmark (1994) Nature 367:116-117.

Also, a ligand is a molecule which serves either as a natural ligand to which said receptor, or an analog 15 thereof, binds, or a molecule which is a functional analog of the natural ligand. The functional analog may be a ligand with structural modifications, or may be a wholly unrelated molecule which has a molecular shape which interacts with the appropriate ligand binding 20 determinants. The ligands may serve as agonists or antagonists, see, e.g., Goodman, et al. (eds) (1990) Goodman & Gilman's: The Pharmacological Bases of Therapeutics, Pergamon Press, New York.

Rational drug design may also be based upon 25 structural studies of the molecular shapes of a receptor or antibody and other effectors or ligands. Effectors may be other proteins which mediate other functions in response to ligand binding, or other proteins which normally interact with the receptor. One means for 30 determining which sites interact with specific other proteins is a physical structure determination, e.g., x-ray crystallography or 2 dimensional NMR techniques. These will provide guidance as to which amino acid residues form molecular contact regions. For a detailed 35 description of protein structural determination, see, e.g., Blundell and Johnson (1976) Protein

Crystallography, Academic Press, New York, which is hereby incorporated herein by reference.

## II. Activities

5       The Toll like receptor proteins will have a number of different biological activities, e.g., in phosphate metabolism, being added to or removed from specific substrates, typically proteins. Such will generally result in modulation of an inflammatory function, other  
10 innate immunity response, or a morphological effect. The DTLR2, 3, 4, 5, 6, 7, 8, 9, or 10 proteins are homologous to other Toll like receptor proteins, but each have structural differences. For example, a human DTLR2 gene coding sequence probably has about 70% identity with the  
15 nucleotide coding sequence of mouse DTLR2. At the amino acid level, there is also likely to be reasonable identity.

The biological activities of the DTLRs will be related to addition or removal of phosphate moieties to substrates, typically in a specific manner, but occasionally in a non specific manner. Substrates may be identified, or conditions for enzymatic activity may be assayed by standard methods, e.g., as described in Hardie, et al. (eds. 1995) The Protein Kinase FactBook  
20 vols. I and II, Academic Press, San Diego, CA; Hanks, et al. (1991) Meth. Enzymol. 200:38-62; Hunter, et al. (1992) Cell 70:375-388; Lewin (1990) Cell 61:743-752;  
25 Pines, et al. (1991) Cold Spring Harbor Symp. Quant. Biol. 56:449-463; and Parker, et al. (1993) Nature 363:736-738.

## III. Nucleic Acids

This invention contemplates use of isolated nucleic acid or fragments, e.g., which encode these or closely related proteins, or fragments thereof, e.g., to encode a corresponding polypeptide, preferably one which is biologically active. In addition, this invention covers  
35

isolated or recombinant DNA which encodes such proteins or polypeptides having characteristic sequences of the respective DTLRs, individually or as a group. Typically, the nucleic acid is capable of hybridizing, under appropriate conditions, with a nucleic acid sequence segment shown in SEQ ID NOs: 3, 5, 25, 9, 11, 15, 17, 31, 21, or 33, but preferably not with a corresponding segment of SEQ ID NO: 1. Said biologically active protein or polypeptide can be a full length protein, or fragment, and will typically have a segment of amino acid sequence highly homologous to one shown in SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34. Further, this invention covers the use of isolated or recombinant nucleic acid, or fragments thereof, which encode proteins having fragments which are equivalent to the DTLR2-10 proteins. The isolated nucleic acids can have the respective regulatory sequences in the 5' and 3' flanks, e.g., promoters, enhancers, poly-A addition signals, and others from the natural gene.

An "isolated" nucleic acid is a nucleic acid, e.g., an RNA, DNA, or a mixed polymer, which is substantially pure, e.g., separated from other components which naturally accompany a native sequence, such as ribosomes, polymerases, and flanking genomic sequences from the originating species. The term embraces a nucleic acid sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates, which are thereby distinguishable from naturally occurring compositions, and chemically synthesized analogs or analogs biologically synthesized by heterologous systems. A substantially pure molecule includes isolated forms of the molecule, either completely or substantially pure.

An isolated nucleic acid will generally be a homogeneous composition of molecules, but will, in some embodiments, contain heterogeneity, preferably minor. This heterogeneity is typically found at the polymer ends

or portions not critical to a desired biological function or activity.

A "recombinant" nucleic acid is typically defined either by its method of production or its structure. In reference to its method of production, e.g., a product made by a process, the process is use of recombinant nucleic acid techniques, e.g., involving human intervention in the nucleotide sequence. Typically this intervention involves in vitro manipulation, although under certain circumstances it may involve more classical animal breeding techniques. Alternatively, it can be a nucleic acid made by generating a sequence comprising fusion of two fragments which are not naturally contiguous to each other, but is meant to exclude products of nature, e.g., naturally occurring mutants as found in their natural state. Thus, for example, products made by transforming cells with any unnaturally occurring vector is encompassed, as are nucleic acids comprising sequence derived using any synthetic oligonucleotide process. Such a process is often done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a restriction enzyme sequence recognition site. Alternatively, the process is performed to join together nucleic acid segments of desired functions to generate a single genetic entity comprising a desired combination of functions not found in the commonly available natural forms, e.g., encoding a fusion protein. Restriction enzyme recognition sites are often the target of such artificial manipulations, but other site specific targets, e.g., promoters, DNA replication sites, regulation sequences, control sequences, or other useful features may be incorporated by design. A similar concept is intended for a recombinant, e.g., fusion, polypeptide. This will include a dimeric repeat. Specifically included are synthetic nucleic acids which, by genetic code redundancy, encode equivalent

polypeptides to fragments of DTLR2-10 and fusions of sequences from various different related molecules, e.g., other IL-1 receptor family members.

A "fragment" in a nucleic acid context is a contiguous segment of at least about 17 nucleotides, generally at least 21 nucleotides, more generally at least 25 nucleotides, ordinarily at least 30 nucleotides, more ordinarily at least 35 nucleotides, often at least 39 nucleotides, more often at least 45 nucleotides, typically at least 50 nucleotides, more typically at least 55 nucleotides, usually at least 60 nucleotides, more usually at least 66 nucleotides, preferably at least 72 nucleotides, more preferably at least 79 nucleotides, and in particularly preferred embodiments will be at least 85 or more nucleotides. Typically, fragments of different genetic sequences can be compared to one another over appropriate length stretches, particularly defined segments such as the domains described below.

A nucleic acid which codes for a DTLR2-10 will be particularly useful to identify genes, mRNA, and cDNA species which code for itself or closely related proteins, as well as DNAs which code for polymorphic, allelic, or other genetic variants, e.g., from different individuals or related species. Preferred probes for such screens are those regions of the interleukin which are conserved between different polymorphic variants or which contain nucleotides which lack specificity, and will preferably be full length or nearly so. In other situations, polymorphic variant specific sequences will be more useful.

This invention further covers recombinant nucleic acid molecules and fragments having a nucleic acid sequence identical to or highly homologous to the isolated DNA set forth herein. In particular, the sequences will often be operably linked to DNA segments which control transcription, translation, and DNA

replication. These additional segments typically assist in expression of the desired nucleic acid segment.

Homologous, or highly identical, nucleic acid sequences, when compared to one another or the sequences shown in SEQ ID NO: 3, 5, 25, 9, 11, 15, 17, 31, 21, or 33 exhibit significant similarity. The standards for homology in nucleic acids are either measures for homology generally used in the art by sequence comparison or based upon hybridization conditions. Comparative hybridization conditions are described in greater detail below.

Substantial identity in the nucleic acid sequence comparison context means either that the segments, or their complementary strands, when compared, are identical when optimally aligned, with appropriate nucleotide insertions or deletions, in at least about 60% of the nucleotides, generally at least 66%, ordinarily at least 71%, often at least 76%, more often at least 80%, usually at least 84%, more usually at least 88%, typically at least 91%, more typically at least about 93%, preferably at least about 95%, more preferably at least about 96 to 98% or more, and in particular embodiments, as high as about 99% or more of the nucleotides, including, e.g., segments encoding structural domains such as the segments described below. Alternatively, substantial identity will exist when the segments will hybridize under selective hybridization conditions, to a strand or its complement, typically using a sequence derived from SEQ ID NO: 3, 5, 25, 9, 11, 15, 17, 31, 21, or 33.

Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, more typically at least about 65%, preferably at least about 75%, and more preferably at least about 90%. See, Kanehisa (1984) Nuc. Acids Res. 12:203-213, which is incorporated herein by reference. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will be

over a stretch of at least about 17 nucleotides, generally at least about 20 nucleotides, ordinarily at least about 24 nucleotides, usually at least about 28 nucleotides, typically at least about 32 nucleotides,  
5 more typically at least about 40 nucleotides, preferably at least about 50 nucleotides, and more preferably at least about 75 to 100 or more nucleotides.

Stringent conditions, in referring to homology in the hybridization context, will be stringent combined  
10 conditions of salt, temperature, organic solvents, and other parameters typically controlled in hybridization reactions. Stringent temperature conditions will usually include temperatures in excess of about 30° C, more usually in excess of about 37° C, typically in excess of about 45° C, more typically in excess of about 55° C, preferably in excess of about 65° C, and more preferably in excess of about 70° C. Stringent salt conditions will ordinarily be less than about 500 mM, usually less than about 400 mM, more usually less than about 300 mM,  
15 typically less than about 200 mM, preferably less than about 100 mM, and more preferably less than about 80 mM, even down to less than about 20 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and  
20 Davidson (1968) J. Mol. Biol. 31:349-370, which is hereby incorporated herein by reference.

Alternatively, for sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison  
30 algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test  
35 sequence(s) relative to the reference sequence, based on the designated program parameters.

Optical alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch 5 (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer 10 Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel et al., *supra*).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments 15 to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) J. Mol. Evol. 35:351-360. The 20 method used is similar to the method described by Higgins and Sharp (1989) CABIOS 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two 25 most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two 30 individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program 35 parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following

parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described Altschul, et al. (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul, et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Nat'l Acad. Sci.

USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ( $P(N)$ ), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

A further indication that two nucleic acid sequences of polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, e.g., where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

The isolated DNA can be readily modified by nucleotide substitutions, nucleotide deletions, nucleotide insertions, and inversions of nucleotide stretches. These modifications result in novel DNA sequences which encode this protein or its derivatives. These modified sequences can be used to produce mutant proteins (muteins) or to enhance the expression of variant species. Enhanced expression may involve gene amplification, increased transcription, increased translation, and other mechanisms. Such mutant DTLR-like derivatives include predetermined or site-specific mutations of the protein or its fragments, including silent mutations using genetic code degeneracy. "Mutant DTLR" as used herein encompasses a polypeptide otherwise falling within the homology definition of the DTLR as set

forth above, but having an amino acid sequence which differs from that of other DTLR-like proteins as found in nature, whether by way of deletion, substitution, or insertion. In particular, "site specific mutant DTLR" 5 encompasses a protein having substantial homology with a protein of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34, and typically shares most of the biological activities or effects of the forms disclosed herein.

Although site specific mutation sites are 10 predetermined, mutants need not be site specific. Mammalian DTLR mutagenesis can be achieved by making amino acid insertions or deletions in the gene, coupled with expression. Substitutions, deletions, insertions, or any combinations may be generated to arrive at a final 15 construct. Insertions include amino- or carboxy-terminal fusions. Random mutagenesis can be conducted at a target codon and the expressed mammalian DTLR mutants can then be screened for the desired activity. Methods for making substitution mutations at predetermined sites 20 in DNA having a known sequence are well known in the art, e.g., by M13 primer mutagenesis. See also Sambrook, et al. (1989) and Ausubel, et al. (1987 and periodic Supplements).

The mutations in the DNA normally should not place 25 coding sequences out of reading frames and preferably will not create complementary regions that could hybridize to produce secondary mRNA structure such as loops or hairpins.

The phosphoramidite method described by Beaucage and 30 Carruthers (1981) Tetra. Letts. 22:1859-1862, will produce suitable synthetic DNA fragments. A double stranded fragment will often be obtained either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding 35 the complementary strand using DNA polymerase with an appropriate primer sequence.

Polymerase chain reaction (PCR) techniques can often be applied in mutagenesis. Alternatively, mutagenesis primers are commonly used methods for generating defined mutations at predetermined sites. See, e.g., Innis, et al. (eds. 1990) PCR Protocols: A Guide to Methods and Applications Academic Press, San Diego, CA; and Dieffenbach and Dveksler (1995; eds.) PCR Primer: A Laboratory Manual Cold Spring Harbor Press, CSH, NY.

10 IV. Proteins, Peptides

As described above, the present invention encompasses primate DTLR2-10, e.g., whose sequences are disclosed in SEQ ID NOS: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34, and described above. Allelic and other variants 15 are also contemplated, including, e.g., fusion proteins combining portions of such sequences with others, including epitope tags and functional domains.

The present invention also provides recombinant proteins, e.g., heterologous fusion proteins using segments from these rodent proteins. A heterologous fusion protein is a fusion of proteins or segments which are naturally not normally fused in the same manner. Thus, the fusion product of a DTLR with an IL-1 receptor 20 is a continuous protein molecule having sequences fused in a typical peptide linkage, typically made as a single 25 translation product and exhibiting properties, e.g., sequence or antigenicity, derived from each source peptide. A similar concept applies to heterologous nucleic acid sequences.

30 In addition, new constructs may be made from combining similar functional or structural domains from other related proteins, e.g., IL-1 receptors or other DTLRs, including species variants. For example, ligand-binding or other segments may be "swapped" between 35 different new fusion polypeptides or fragments. See, e.g., Cunningham, et al. (1989) Science 243:1330-1336; and O'Dowd, et al. (1988) J. Biol. Chem. 263:15985-15992,

each of which is incorporated herein by reference. Thus, new chimeric polypeptides exhibiting new combinations of specificities will result from the functional linkage of receptor-binding specificities. For example, the ligand 5 binding domains from other related receptor molecules may be added or substituted for other domains of this or related proteins. The resulting protein will often have hybrid function and properties. For example, a fusion protein may include a targetting domain which may serve 10 to provide sequestering of the fusion protein to a particular subcellular organelle.

Candidate fusion partners and sequences can be selected from various sequence data bases, e.g., GenBank, c/o IntelliGenetics, Mountain View, CA; and BCG, 15 University of Wisconsin Biotechnology Computing Group, Madison, WI, which are each incorporated herein by reference.

The present invention particularly provides muteins which bind Toll ligands, and/or which are affected in 20 signal transduction. Structural alignment of human DTLR1-10 with other members of the IL-1 family show conserved features/residues. See, e.g., Figure 3A. Alignment of the human DTLR sequences with other members of the IL-1 family indicates various structural and 25 functionally shared features. See also, Bazan, et al. (1996) Nature 379:591; Lodi, et al. (1994) Science 263:1762-1766; Sayle and Milner-White (1995) TIBS 20:374-376; and Gronenberg, et al. (1991) Protein Engineering 4:263-269.

30 The IL-1 $\alpha$  and IL-1 $\beta$  ligands bind an IL-1 receptor type I as the primary receptor and this complex then forms a high affinity receptor complex with the IL-1 receptor type III. Such receptor subunits are probably shared with the new IL-1 family members.

35 Similar variations in other species counterparts of DTLR2-10 sequences, e.g., in the corresponding regions, should provide similar interactions with ligand or

substrate. Substitutions with either mouse sequences or human sequences are particularly preferred. Conversely, conservative substitutions away from the ligand binding interaction regions will probably preserve most signaling activities.

"Derivatives" of the primate DTLR2-10 include amino acid sequence mutants, glycosylation variants, metabolic derivatives and covalent or aggregative conjugates with other chemical moieties. Covalent derivatives can be prepared by linkage of functionalities to groups which are found in the DTLR amino acid side chains or at the N- or C- termini, e.g., by means which are well known in the art. These derivatives can include, without limitation, aliphatic esters or amides of the carboxyl terminus, or of residues containing carboxyl side chains, O-acyl derivatives of hydroxyl group-containing residues, and N-acyl derivatives of the amino terminal amino acid or amino-group containing residues, e.g., lysine or arginine. Acyl groups are selected from the group of alkyl-moieties including C3 to C18 normal alkyl, thereby forming alkanoyl aroyl species.

In particular, glycosylation alterations are included, e.g., made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing, or in further processing steps. Particularly preferred means for accomplishing this are by exposing the polypeptide to glycosylating enzymes derived from cells which normally provide such processing, e.g., mammalian glycosylation enzymes. Deglycosylation enzymes are also contemplated. Also embraced are versions of the same primary amino acid sequence which have other minor modifications, including phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine.

A major group of derivatives are covalent conjugates of the receptors or fragments thereof with other proteins of polypeptides. These derivatives can be synthesized in

recombinant culture such as N- or C-terminal fusions or by the use of agents known in the art for their usefulness in cross-linking proteins through reactive side groups. Preferred derivatization sites with 5 cross-linking agents are at free amino groups, carbohydrate moieties, and cysteine residues.

Fusion polypeptides between the receptors and other homologous or heterologous proteins are also provided. Homologous polypeptides may be fusions between different 10 receptors, resulting in, for instance, a hybrid protein exhibiting binding specificity for multiple different Toll ligands, or a receptor which may have broadened or weakened specificity of substrate effect. Likewise, heterologous fusions may be constructed which would 15 exhibit a combination of properties or activities of the derivative proteins. Typical examples are fusions of a reporter polypeptide, e.g., luciferase, with a segment or domain of a receptor, e.g., a ligand-binding segment, so that the presence or location of a desired ligand may be 20 easily determined. See, e.g., Dull, et al., U.S. Patent No. 4,859,609, which is hereby incorporated herein by reference. Other gene fusion partners include glutathione-S-transferase (GST), bacterial  $\beta$ -galactosidase, trpE, Protein A,  $\beta$ -lactamase, alpha 25 amylase, alcohol dehydrogenase, and yeast alpha mating factor. See, e.g., Godowski, et al. (1988) Science 241:812-816.

The phosphoramidite method described by Beaucage and Carruthers (1981) Tetra. Letts. 22:1859-1862, will 30 produce suitable synthetic DNA fragments. A double stranded fragment will often be obtained either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an 35 appropriate primer sequence.

Such polypeptides may also have amino acid residues which have been chemically modified by phosphorylation,

sulfonation, biotinylation, or the addition or removal of other moieties, particularly those which have molecular shapes similar to phosphate groups. In some embodiments, the modifications will be useful labeling reagents, or 5 serve as purification targets, e.g., affinity ligands.

Fusion proteins will typically be made by either recombinant nucleic acid methods or by synthetic polypeptide methods. Techniques for nucleic acid manipulation and expression are described generally, for 10 example, in Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.), Vols. 1-3, Cold Spring Harbor Laboratory, and Ausubel, et al. (eds. 1987 and periodic supplements) Current Protocols in Molecular Biology, Greene/Wiley, New York, which are each incorporated 15 herein by reference. Techniques for synthesis of polypeptides are described, for example, in Merrifield (1963) J. Amer. Chem. Soc. 85:2149-2156; Merrifield (1986) Science 232: 341-347; and Atherton, et al. (1989) Solid Phase Peptide Synthesis: A Practical Approach, IRL 20 Press, Oxford; each of which is incorporated herein by reference. See also Dawson, et al. (1994) Science 266:776-779 for methods to make larger polypeptides.

This invention also contemplates the use of derivatives of a DTLR2-10 other than variations in amino 25 acid sequence or glycosylation. Such derivatives may involve covalent or aggregative association with chemical moieties. These derivatives generally fall into three classes: (1) salts, (2) side chain and terminal residue covalent modifications, and (3) adsorption complexes, for 30 example with cell membranes. Such covalent or aggregative derivatives are useful as immunogens, as reagents in immunoassays, or in purification methods such as for affinity purification of a receptor or other binding molecule, e.g., an antibody. For example, a Toll 35 ligand can be immobilized by covalent bonding to a solid support such as cyanogen bromide-activated Sepharose, by methods which are well known in the art, or adsorbed onto

polyolefin surfaces, with or without glutaraldehyde cross-linking, for use in the assay or purification of a DTLR receptor, antibodies, or other similar molecules. The ligand can also be labeled with a detectable group,  
5 for example radioiodinated by the chloramine T procedure, covalently bound to rare earth chelates, or conjugated to another fluorescent moiety for use in diagnostic assays.

A DTLR of this invention can be used as an immunogen for the production of antisera or antibodies specific,  
10 e.g., capable of distinguishing between other IL-1 receptor family members, for the DTLR or various fragments thereof. The purified DTLR can be used to screen monoclonal antibodies or antigen-binding fragments prepared by immunization with various forms of impure  
15 preparations containing the protein. In particular, the term "antibodies" also encompasses antigen binding fragments of natural antibodies, e.g., Fab, Fab2, Fv, etc. The purified DTLR can also be used as a reagent to detect antibodies generated in response to the presence  
20 of elevated levels of expression, or immunological disorders which lead to antibody production to the endogenous receptor. Additionally, DTLR fragments may also serve as immunogens to produce the antibodies of the present invention, as described immediately below. For  
25 example, this invention contemplates antibodies having binding affinity to or being raised against the amino acid sequences shown in SEQ ID NOS: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34, fragments thereof, or various homologous peptides. In particular, this invention  
30 contemplates antibodies having binding affinity to, or having been raised against, specific fragments which are predicted to be, or actually are, exposed at the exterior protein surface of the native DTLR.

The blocking of physiological response to the  
35 receptor ligands may result from the inhibition of binding of the ligand to the receptor, likely through competitive inhibition. Thus, in vitro assays of the

present invention will often use antibodies or antigen binding segments of these antibodies, or fragments attached to solid phase substrates. These assays will also allow for the diagnostic determination of the 5 effects of either ligand binding region mutations and modifications, or other mutations and modifications, e.g., which affect signaling or enzymatic function.

This invention also contemplates the use of competitive drug screening assays, e.g., where 10 neutralizing antibodies to the receptor or fragments compete with a test compound for binding to a ligand or other antibody. In this manner, the neutralizing antibodies or fragments can be used to detect the presence of a polypeptide which shares one or more 15 binding sites to a receptor and can also be used to occupy binding sites on a receptor that might otherwise bind a ligand.

#### V. Making Nucleic Acids and Protein

20 DNA which encodes the protein or fragments thereof can be obtained by chemical synthesis, screening cDNA libraries, or by screening genomic libraries prepared from a wide variety of cell lines or tissue samples. Natural sequences can be isolated using standard methods 25 and the sequences provided herein. Other species counterparts can be identified by hybridization techniques, or by various PCR techniques, combined with or by searching in sequence databases, e.g., GenBank.

This DNA can be expressed in a wide variety of host 30 cells for the synthesis of a full-length receptor or fragments which can in turn, for example, be used to generate polyclonal or monoclonal antibodies; for binding studies; for construction and expression of modified ligand binding or kinase/phosphatase domains; and for 35 structure/function studies. Variants or fragments can be expressed in host cells that are transformed or transfected with appropriate expression vectors. These

molecules can be substantially free of protein or cellular contaminants, other than those derived from the recombinant host, and therefore are particularly useful in pharmaceutical compositions when combined with a 5 pharmaceutically acceptable carrier and/or diluent. The protein, or portions thereof, may be expressed as fusions with other proteins.

Expression vectors are typically self-replicating DNA or RNA constructs containing the desired receptor 10 gene or its fragments, usually operably linked to suitable genetic control elements that are recognized in a suitable host cell. These control elements are capable of effecting expression within a suitable host. The specific type of control elements necessary to effect 15 expression will depend upon the eventual host cell used. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system, and typically include a transcriptional promoter, an optional operator to control 20 the onset of transcription, transcription enhancers to elevate the level of mRNA expression, a sequence that encodes a suitable ribosome binding site, and sequences that terminate transcription and translation. Expression vectors also usually contain an origin of replication 25 that allows the vector to replicate independently of the host cell.

The vectors of this invention include those which contain DNA which encodes a protein, as described, or a fragment thereof encoding a biologically active 30 equivalent polypeptide. The DNA can be under the control of a viral promoter and can encode a selection marker. This invention further contemplates use of such expression vectors which are capable of expressing eukaryotic cDNA coding for such a protein in a 35 prokaryotic or eukaryotic host, where the vector is compatible with the host and where the eukaryotic cDNA coding for the receptor is inserted into the vector such

that growth of the host containing the vector expresses the cDNA in question. Usually, expression vectors are designed for stable replication in their host cells or for amplification to greatly increase the total number of 5 copies of the desirable gene per cell. It is not always necessary to require that an expression vector replicate in a host cell, e.g., it is possible to effect transient expression of the protein or its fragments in various hosts using vectors that do not contain a replication 10 origin that is recognized by the host cell. It is also possible to use vectors that cause integration of the protein encoding portion or its fragments into the host DNA by recombination.

Vectors, as used herein, comprise plasmids, viruses, 15 bacteriophage, integratable DNA fragments, and other vehicles which enable the integration of DNA fragments into the genome of the host. Expression vectors are specialized vectors which contain genetic control elements that effect expression of operably linked genes. 20 Plasmids are the most commonly used form of vector but all other forms of vectors which serve an equivalent function and which are, or become, known in the art are suitable for use herein. See, e.g., Pouwels, et al. (1985 and Supplements) Cloning Vectors: A Laboratory 25 Manual, Elsevier, N.Y., and Rodriguez, et al. (eds) Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworth, Boston, 1988, which are incorporated herein by reference.

Transformed cells are cells, preferably mammalian, 30 that have been transformed or transfected with receptor vectors constructed using recombinant DNA techniques. Transformed host cells usually express the desired protein or its fragments, but for purposes of cloning, amplifying, and manipulating its DNA, do not need to 35 express the subject protein. This invention further contemplates culturing transformed cells in a nutrient medium, thus permitting the receptor to accumulate in the

cell membrane. The protein can be recovered, either from the culture or, in certain instances, from the culture medium.

For purposes of this invention, nucleic sequences are operably linked when they are functionally related to each other. For example, DNA for a presequence or secretory leader is operably linked to a polypeptide if it is expressed as a preprotein or participates in directing the polypeptide to the cell membrane or in secretion of the polypeptide. A promoter is operably linked to a coding sequence if it controls the transcription of the polypeptide; a ribosome binding site is operably linked to a coding sequence if it is positioned to permit translation. Usually, operably linked means contiguous and in reading frame, however, certain genetic elements such as repressor genes are not contiguously linked but still bind to operator sequences that in turn control expression.

Suitable host cells include prokaryotes, lower eukaryotes, and higher eukaryotes. Prokaryotes include both gram negative and gram positive organisms, e.g., E. coli and B. subtilis. Lower eukaryotes include yeasts, e.g., S. cerevisiae and Pichia, and species of the genus Dictyostelium. Higher eukaryotes include established tissue culture cell lines from animal cells, both of non-mammalian origin, e.g., insect cells, and birds, and of mammalian origin, e.g., human, primates, and rodents.

Prokaryotic host-vector systems include a wide variety of vectors for many different species. As used herein, E. coli and its vectors will be used generically to include equivalent vectors used in other prokaryotes. A representative vector for amplifying DNA is pBR322 or many of its derivatives. Vectors that can be used to express the receptor or its fragments include, but are not limited to, such vectors as those containing the lac promoter (pUC-series); trp promoter (pBR322-trp); I<sub>pp</sub> promoter (the pIN-series); lambda-pP or pR promoters

(pOTS); or hybrid promoters such as ptac (pDR540). See Brosius, et al. (1988) "Expression Vectors Employing Lambda-, trp-, lac-, and Ipp-derived Promoters", in Vectors: A Survey of Molecular Cloning Vectors and Their Uses, (eds. Rodriguez and Denhardt), Butterworth, Boston, Chapter 10, pp. 205-236, which is incorporated herein by reference.

Lower eukaryotes, e.g., yeasts and Dictyostelium, may be transformed with DTLR sequence containing vectors. For purposes of this invention, the most common lower eukaryotic host is the baker's yeast, Saccharomyces cerevisiae. It will be used to generically represent lower eukaryotes although a number of other strains and species are also available. Yeast vectors typically consist of a replication origin (unless of the integrating type), a selection gene, a promoter, DNA encoding the receptor or its fragments, and sequences for translation termination, polyadenylation, and transcription termination. Suitable expression vectors for yeast include such constitutive promoters as 3-phosphoglycerate kinase and various other glycolytic enzyme gene promoters or such inducible promoters as the alcohol dehydrogenase 2 promoter or metallothioneine promoter. Suitable vectors include derivatives of the following types: self-replicating low copy number (such as the YRp-series), self-replicating high copy number (such as the YEp-series); integrating types (such as the YIp-series), or mini-chromosomes (such as the YCp-series).

Higher eukaryotic tissue culture cells are normally the preferred host cells for expression of the functionally active interleukin protein. In principle, any higher eukaryotic tissue culture cell line is workable, e.g., insect baculovirus expression systems, whether from an invertebrate or vertebrate source. However, mammalian cells are preferred. Transformation or transfection and propagation of such cells has become

a routine procedure. Examples of useful cell lines include HeLa cells, Chinese hamster ovary (CHO) cell lines, baby rat kidney (BRK) cell lines, insect cell lines, bird cell lines, and monkey (COS) cell lines.

5 Expression vectors for such cell lines usually include an origin of replication, a promoter, a translation initiation site, RNA splice sites (if genomic DNA is used), a polyadenylation site, and a transcription termination site. These vectors also usually contain a  
10 selection gene or amplification gene. Suitable expression vectors may be plasmids, viruses, or retroviruses carrying promoters derived, e.g., from such sources as from adenovirus, SV40, parvoviruses, vaccinia virus, or cytomegalovirus. Representative examples of  
15 suitable expression vectors include pCDNA1; pCD, see Okayama, et al. (1985) Mol. Cell Biol. 5:1136-1142; pMC1neo PolyA, see Thomas, et al. (1987) Cell 51:503-512; and a baculovirus vector such as pAC 373 or pAC 610.

For secreted proteins, an open reading frame usually  
20 encodes a polypeptide that consists of a mature or secreted product covalently linked at its N-terminus to a signal peptide. The signal peptide is cleaved prior to secretion of the mature, or active, polypeptide. The cleavage site can be predicted with a high degree of  
25 accuracy from empirical rules, e.g., von-Heijne (1986) Nucleic Acids Research 14:4683-4690, and the precise amino acid composition of the signal peptide does not appear to be critical to its function, e.g., Randall, et al. (1989) Science 243:1156-1159; Kaiser et al. (1987)  
30 Science 235:312-317.

It will often be desired to express these polypeptides in a system which provides a specific or defined glycosylation pattern. In this case, the usual pattern will be that provided naturally by the expression  
35 system. However, the pattern will be modifiable by exposing the polypeptide, e.g., an unglycosylated form, to appropriate glycosylating proteins introduced into a

heterologous expression system. For example, the receptor gene may be co-transformed with one or more genes encoding mammalian or other glycosylating enzymes. Using this approach, certain mammalian glycosylation patterns will be achievable in prokaryote or other cells.

The source of DTLR can be a eukaryotic or prokaryotic host expressing recombinant DTLR, such as is described above. The source can also be a cell line such as mouse Swiss 3T3 fibroblasts, but other mammalian cell lines are also contemplated by this invention, with the preferred cell line being from the human species.

Now that the sequences are known, the primate DTLRs, fragments, or derivatives thereof can be prepared by conventional processes for synthesizing peptides. These include processes such as are described in Stewart and Young (1984) Solid Phase Peptide Synthesis, Pierce Chemical Co., Rockford, IL; Bodanszky and Bodanszky (1984) The Practice of Peptide Synthesis, Springer-Verlag, New York; and Bodanszky (1984) The Principles of Peptide Synthesis, Springer-Verlag, New York; all of each which are incorporated herein by reference. For example, an azide process, an acid chloride process, an acid anhydride process, a mixed anhydride process, an active ester process (e.g., p-nitrophenyl ester, N-hydroxysuccinimide ester, or cyanomethyl ester), a carbodiimidazole process, an oxidative-reductive process, or a dicyclohexylcarbodiimide (DCCD)/additive process can be used. Solid phase and solution phase syntheses are both applicable to the foregoing processes. Similar techniques can be used with partial DTLR sequences.

The DTLR proteins, fragments, or derivatives are suitably prepared in accordance with the above processes as typically employed in peptide synthesis, generally either by a so-called stepwise process which comprises condensing an amino acid to the terminal amino acid, one by one in sequence, or by coupling peptide fragments to

the terminal amino acid. Amino groups that are not being used in the coupling reaction typically must be protected to prevent coupling at an incorrect location.

If a solid phase synthesis is adopted, the  
5 C-terminal amino acid is bound to an insoluble carrier or support through its carboxyl group. The insoluble carrier is not particularly limited as long as it has a binding capability to a reactive carboxyl group.  
Examples of such insoluble carriers include halomethyl  
10 resins, such as chloromethyl resin or bromomethyl resin, hydroxymethyl resins, phenol resins, tert-alkyloxycarbonylhydrazidated resins, and the like.

An amino group-protected amino acid is bound in sequence through condensation of its activated carboxyl  
15 group and the reactive amino group of the previously formed peptide or chain, to synthesize the peptide step by step. After synthesizing the complete sequence, the peptide is split off from the insoluble carrier to produce the peptide. This solid-phase approach is  
20 generally described by Merrifield, et al. (1963) in J. Am. Chem. Soc. 85:2149-2156, which is incorporated herein by reference.

The prepared protein and fragments thereof can be isolated and purified from the reaction mixture by means  
25 of peptide separation, for example, by extraction, precipitation, electrophoresis, various forms of chromatography, and the like. The receptors of this invention can be obtained in varying degrees of purity depending upon desired uses. Purification can be  
30 accomplished by use of the protein purification techniques disclosed herein, see below, or by the use of the antibodies herein described in methods of immunoabsorbant affinity chromatography. This immunoabsorbant affinity chromatography is carried out by  
35 first linking the antibodies to a solid support and then contacting the linked antibodies with solubilized lysates of appropriate cells, lysates of other cells expressing

the receptor, or lysates or supernatants of cells producing the protein as a result of DNA techniques, see below.

Generally, the purified protein will be at least 5 about 40% pure, ordinarily at least about 50% pure, usually at least about 60% pure, typically at least about 70% pure, more typically at least about 80% pure, preferable at least about 90% pure and more preferably at least about 95% pure, and in particular embodiments, 97%-10 99% or more. Purity will usually be on a weight basis, but can also be on a molar basis. Different assays will be applied as appropriate.

#### VI. Antibodies

15       Antibodies can be raised to the various mammalian, e.g., primate DTLR proteins and fragments thereof, both in naturally occurring native forms and in their recombinant forms, the difference being that antibodies to the active receptor are more likely to recognize 20 epitopes which are only present in the native conformations. Denatured antigen detection can also be useful in, e.g., Western analysis. Anti-idiotypic antibodies are also contemplated, which would be useful as agonists or antagonists of a natural receptor or an 25 antibody.

Antibodies, including binding fragments and single chain versions, against predetermined fragments of the protein can be raised by immunization of animals with conjugates of the fragments with immunogenic proteins. 30 Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective protein, or screened for agonistic or antagonistic activity. These monoclonal antibodies will usually bind with at least a  $K_D$  of about 35 1 mM, more usually at least about 300  $\mu\text{M}$ , typically at least about 100 $\mu\text{M}$ , more typically at least about 30  $\mu\text{M}$ ,

preferably at least about 10  $\mu\text{M}$ , and more preferably at least about 3  $\mu\text{M}$  or better.

The antibodies, including antigen binding fragments, of this invention can have significant diagnostic or therapeutic value. They can be potent antagonists that bind to the receptor and inhibit binding to ligand or inhibit the ability of the receptor to elicit a biological response, e.g., act on its substrate. They also can be useful as non-neutralizing antibodies and can be coupled to toxins or radionuclides to bind producing cells, or cells localized to the source of the interleukin. Further, these antibodies can be conjugated to drugs or other therapeutic agents, either directly or indirectly by means of a linker.

The antibodies of this invention can also be useful in diagnostic applications. As capture or non-neutralizing antibodies, they might bind to the receptor without inhibiting ligand or substrate binding. As neutralizing antibodies, they can be useful in competitive binding assays. They will also be useful in detecting or quantifying ligand. They may be used as reagents for Western blot analysis, or for immunoprecipitation or immunopurification of the respective protein.

Protein fragments may be joined to other materials, particularly polypeptides, as fused or covalently joined polypeptides to be used as immunogens. Mammalian DTLR and its fragments may be fused or covalently linked to a variety of immunogens, such as keyhole limpet hemocyanin, bovine serum albumin, tetanus toxoid, etc. See Microbiology, Hoeber Medical Division, Harper and Row, 1969; Landsteiner (1962) Specificity of Serological Reactions, Dover Publications, New York; and Williams, et al. (1967) Methods in Immunology and Immunochemistry, Vol. 1, Academic Press, New York; each of which are incorporated herein by reference, for descriptions of methods of preparing polyclonal antisera. A typical

method involves hyperimmunization of an animal with an antigen. The blood of the animal is then collected shortly after the repeated immunizations and the gamma globulin is isolated.

5        In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites, et al. (eds) Basic and  
10 Clinical Immunology (4th ed.), Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed) Academic Press, New York;  
15 and particularly in Kohler and Milstein (1975) in Nature 256: 495-497, which discusses one method of generating monoclonal antibodies. Each of these references is incorporated herein by reference. Summarized briefly, this method involves injecting an animal with an  
20 immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells. The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro. The population of hybridomas is then screened to isolate individual clones,  
25 each of which secrete a single antibody species to the immunogen. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B cells from the immune animal generated in response to a specific site recognized on the immunogenic  
30 substance.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides or alternatively to selection of libraries of antibodies in phage or similar vectors. See, Huse, et al. (1989) "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda," Science 246:1275-1281; and Ward, et al. (1989) Nature 341:544-

546, each of which is hereby incorporated herein by reference. The polypeptides and antibodies of the present invention may be used with or without modification, including chimeric or humanized antibodies.

5 Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific 10 and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 15 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant or chimeric immunoglobulins may be produced, see Cabilly, U.S. Patent No. 4,816,567; or made in transgenic mice, see Mendez, et al. (1997) Nature Genetics 15:146-156. These references 20 are incorporated herein by reference.

The antibodies of this invention can also be used for affinity chromatography in isolating the DTLRs. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, 25 Sephadex, or the like, where a cell lysate may be passed through the column, the column washed, followed by increasing concentrations of a mild denaturant, whereby the purified protein will be released. The protein may be used to purify antibody.

30 The antibodies may also be used to screen expression libraries for particular expression products. Usually the antibodies used in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

35 Antibodies raised against a DTLR will also be used to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological

conditions related to expression of the protein or cells which express the protein. They also will be useful as agonists or antagonists of the ligand, which may be competitive inhibitors or substitutes for naturally occurring ligands.

A DTLR protein that specifically binds to or that is specifically immunoreactive with an antibody generated against a defined immunogen, such as an immunogen consisting of the amino acid sequence of SEQ ID NO: 4, 6, 10, 26, 10, 12, 16, 18, 32, 22 or 34, is typically determined in an immunoassay. The immunoassay typically uses a polyclonal antiserum which was raised, e.g., to a protein of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34. This antiserum is selected to have low crossreactivity 15 against other IL-1R family members, e.g., DTLR1, preferably from the same species, and any such crossreactivity is removed by immunoabsorption prior to use in the immunoassay.

In order to produce antisera for use in an 20 immunoassay, the protein of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34, or a combination thereof, is isolated as described herein. For example, recombinant protein may be produced in a mammalian cell line. An appropriate host, e.g., an inbred strain of mice such as 25 balb/c, is immunized with the selected protein, typically using a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see Harlow and Lane, *supra*). Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Polyclonal 30 sera are collected and titered against the immunogen protein in an immunoassay, e.g., a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of  $10^4$  or 35 greater are selected and tested for their cross reactivity against other IL-1R family members, e.g., mouse DTLRs or human DTLR1, using a competitive binding

immunoassay such as the one described in Harlow and Lane, supra, at pages 570-573. Preferably at least two DTLR family members are used in this determination in conjunction with either or some of the human DTLR2-10.

- 5 These IL-1R family members can be produced as recombinant proteins and isolated using standard molecular biology and protein chemistry techniques as described herein.

Immunoassays in the competitive binding format can be used for the crossreactivity determinations. For 10 example, the proteins of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34, or various fragments thereof, can be immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins 15 to compete with the binding of the antisera to the immobilized protein is compared to the protein of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 and/or 34. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera 20 with less than 10% crossreactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorbtion with the above-listed proteins.

25 The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein to the immunogen protein (e.g., the IL-1R like protein of SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 and/or 34). In order to make this 30 comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than twice the amount 35 of the protein of the selected protein or proteins that is required, then the second protein is said to

specifically bind to an antibody generated to the immunogen.

It is understood that these DTLR proteins are members of a family of homologous proteins that comprise at least 10 so far identified genes. For a particular gene product, such as the DTLR2-10, the term refers not only to the amino acid sequences disclosed herein, but also to other proteins that are allelic, non-allelic or species variants. It also understood that the terms include nonnatural mutations introduced by deliberate mutation using conventional recombinant technology such as single site mutation, or by excising short sections of DNA encoding the respective proteins, or by substituting new amino acids, or adding new amino acids. Such minor alterations must substantially maintain the immunoidentity of the original molecule and/or its biological activity. Thus, these alterations include proteins that are specifically immunoreactive with a designated naturally occurring IL-1R related protein, for example, the DTLR proteins shown in SEQ ID NO: 4, 6, 26, 10, 12, 16, 18, 32, 22 or 34. The biological properties of the altered proteins can be determined by expressing the protein in an appropriate cell line and measuring the appropriate effect upon lymphocytes. Particular protein modifications considered minor would include conservative substitution of amino acids with similar chemical properties, as described above for the IL-1R family as a whole. By aligning a protein optimally with the protein of DTLR2-10 and by using the conventional immunoassays described herein to determine immunoidentity, one can determine the protein compositions of the invention.

## VII. Kits and quantitation

Both naturally occurring and recombinant forms of the IL-1R like molecules of this invention are particularly useful in kits and assay methods. For example, these methods would also be applied to screening

for binding activity, e.g., ligands for these proteins. Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds per year. See, e.g., a BIOMEK 5 automated workstation, Beckman Instruments, Palo Alto, California, and Fodor, et al. (1991) Science 251:767-773, which is incorporated herein by reference. The latter describes means for testing binding by a plurality of defined polymers synthesized on a solid substrate. The 10 development of suitable assays to screen for a ligand or agonist/antagonist homologous proteins can be greatly facilitated by the availability of large amounts of purified, soluble DTLRs in an active state such as is provided by this invention.

15 Purified DTLR can be coated directly onto plates for use in the aforementioned ligand screening techniques. However, non-neutralizing antibodies to these proteins can be used as capture antibodies to immobilize the respective receptor on the solid phase, useful, e.g., in 20 diagnostic uses.

This invention also contemplates use of DTLR2-10, fragments thereof, peptides, and their fusion products in a variety of diagnostic kits and methods for detecting the presence of the protein or its ligand. 25 Alternatively, or additionally, antibodies against the molecules may be incorporated into the kits and methods. Typically the kit will have a compartment containing either a defined DTLR peptide or gene segment or a reagent which recognizes one or the other. Typically, 30 recognition reagents, in the case of peptide, would be a receptor or antibody, or in the case of a gene segment, would usually be a hybridization probe.

A preferred kit for determining the concentration of, e.g., DTLR4, a sample would typically comprise a 35 labeled compound, e.g., ligand or antibody, having known binding affinity for DTLR4, a source of DTLR4 (naturally occurring or recombinant) as a positive control, and a

means for separating the bound from free labeled compound, for example a solid phase for immobilizing the DTLR4 in the test sample. Compartments containing reagents, and instructions, will normally be provided.

- 5       Antibodies, including antigen binding fragments, specific for mammalian DTLR or a peptide fragment, or receptor fragments are useful in diagnostic applications to detect the presence of elevated levels of ligand and/or its fragments. Diagnostic assays may be
- 10      homogeneous (without a separation step between free reagent and antibody-antigen complex) or heterogeneous (with a separation step). Various commercial assays exist, such as radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA),
- 15      enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA) and the like. For example, unlabeled antibodies can be employed by using a second antibody which is labeled and which recognizes the antibody to DTLR4 or to a particular
- 20      fragment thereof. These assays have also been extensively discussed in the literature. See, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH., and Coligan (Ed.) (1991) and periodic supplements, Current Protocols In Immunology Greene/Wiley, New York.
- 25      Anti-idiotypic antibodies may have similar use to serve as agonists or antagonists of DTLR4. These should be useful as therapeutic reagents under appropriate circumstances.

- 30      Frequently, the reagents for diagnostic assays are supplied in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled antibody, or labeled ligand is provided. This is usually in conjunction with other
- 35      additives, such as buffers, stabilizers, materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also

contain instructions for proper use and disposal of the contents after use. Typically the kit has compartments for each useful reagent, and will contain instructions for proper use and disposal of reagents. Desirably, the 5 reagents are provided as a dry lyophilized powder, where the reagents may be reconstituted in an aqueous medium having appropriate concentrations for performing the assay.

The aforementioned constituents of the diagnostic assays 10 may be used without modification or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, a test compound, DTLR, 15 or antibodies thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups: radiolabels such as  $^{125}\text{I}$ , enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 20 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Both of the patents are incorporated herein by reference. Possibilities for indirect labeling include biotinylation of one constituent followed by 25 binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating the bound from the free ligand, or alternatively the bound from the free test compound. The DTLR can be immobilized 30 on various matrixes followed by washing. Suitable matrices include plastic such as an ELISA plate, filters, and beads. Methods of immobilizing the receptor to a matrix include, without limitation, direct adhesion to plastic, use of a capture antibody, chemical coupling, 35 and biotin-avidin. The last step in this approach involves the precipitation of antibody/antigen complex by any of several methods including those utilizing, e.g.,

an organic solvent such as polyethylene glycol or a salt such as ammonium sulfate. Other suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in  
5 Rattle, et al. (1984) Clin. Chem. 30(9):1457-1461, and the double antibody magnetic particle separation as described in U.S. Pat. No. 4,659,678, each of which is incorporated herein by reference.

The methods for linking protein or fragments to  
10 various labels have been extensively reported in the literature and do not require detailed discussion here. Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of  
15 thioethers by reaction of a mercapto group with an activated halogen such as chloroacetyl, or an activated olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Another diagnostic aspect of this invention involves  
20 use of oligonucleotide or polynucleotide sequences taken from the sequence of a DTLR. These sequences can be used as probes for detecting levels of the respective DTLR in patients suspected of having an immunochemical disorder. The preparation of both RNA and DNA nucleotide sequences,  
25 the labeling of the sequences, and the preferred size of the sequences has received ample description and discussion in the literature. Normally an oligonucleotide probe should have at least about 14 nucleotides, usually at least about 18 nucleotides, and  
30 the polynucleotide probes may be up to several kilobases. Various labels may be employed, most commonly radionuclides, particularly  $^{32}\text{P}$ . However, other techniques may also be employed, such as using biotin modified nucleotides for introduction into a  
35 polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides,

fluorescers, enzymes, or the like. Alternatively, antibodies may be employed which can recognize specific duplexes, including DNA duplexes, RNA duplexes, DNA-RNA hybrid duplexes, or DNA-protein duplexes. The antibodies 5 in turn may be labeled and the assay carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. The use of probes to the novel anti-sense RNA may be carried out in 10 any conventional techniques such as nucleic acid hybridization, plus and minus screening, recombinational probing, hybrid released translation (HRT), and hybrid arrested translation (HART). This also includes amplification techniques such as polymerase chain 15 reaction (PCR).

Diagnostic kits which also test for the qualitative or quantitative presence of other markers are also contemplated. Diagnosis or prognosis may depend on the combination of multiple indications used as markers. 20 Thus, kits may test for combinations of markers. See, e.g., Viallet, et al. (1989) Progress in Growth Factor Res. 1:89-97.

#### VIII. Therapeutic Utility

This invention provides reagents with significant therapeutic value. The DTLRs (naturally occurring or recombinant), fragments thereof, mutein receptors, and antibodies, along with compounds identified as having binding affinity to the receptors or antibodies, should 25 be useful in the treatment of conditions exhibiting abnormal expression of the receptors of their ligands. Such abnormality will typically be manifested by immunological disorders. Additionally, this invention should provide therapeutic value in various diseases or 30 disorders associated with abnormal expression or abnormal triggering of response to the ligand. The Toll ligands 35 have been suggested to be involved in morphologic

development, e.g., dorso-ventral polarity determination, and immune responses, particularly the primitive innate responses. See, e.g., Sun, et al. (1991) Eur. J. Biochem. 196:247-254; Hultmark (1994) Nature 367:116-117.

- 5        Recombinant DTLRs, muteins, agonist or antagonist antibodies thereto, or antibodies can be purified and then administered to a patient. These reagents can be combined for therapeutic use with additional active ingredients, e.g., in conventional pharmaceutically acceptable carriers or diluents, along with physiologically innocuous stabilizers and excipients. These combinations can be sterile, e.g., filtered, and placed into dosage forms as by lyophilization in dosage vials or storage in stabilized aqueous preparations.
- 10      This invention also contemplates use of antibodies or binding fragments thereof which are not complement binding.
- 15      Ligand screening using DTLR or fragments thereof can be performed to identify molecules having binding affinity to the receptors. Subsequent biological assays can then be utilized to determine if a putative ligand can provide competitive binding, which can block intrinsic stimulating activity. Receptor fragments can be used as a blocker or antagonist in that it blocks the activity of ligand. Likewise, a compound having intrinsic stimulating activity can activate the receptor and is thus an agonist in that it simulates the activity of ligand, e.g., inducing signaling. This invention further contemplates the therapeutic use of antibodies to DTLRs as antagonists.

20      The quantities of reagents necessary for effective therapy will depend upon many different factors, including means of administration, target site, physiological state of the patient, and other medicants administered. Thus, treatment dosages should be titrated to optimize safety and efficacy. Typically, dosages used in vitro may provide useful guidance in the amounts

25      30      35

useful for in situ administration of these reagents. Animal testing of effective doses for treatment of particular disorders will provide further predictive indication of human dosage. Various considerations are 5 described, e.g., in Gilman, et al. (eds) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences, (current edition), Mack Publishing Co., Easton, Penn.; each of which is hereby incorporated herein by reference. Methods for administration are discussed 10 therein and below, e.g., for oral, intravenous, intraperitoneal, or intramuscular administration, transdermal diffusion, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, 15 and other compounds described, e.g., in the Merck Index, Merck & Co., Rahway, New Jersey. Because of the likely high affinity binding, or turnover numbers, between a putative ligand and its receptors, low dosages of these reagents would be initially expected to be effective. 20 And the signaling pathway suggests extremely low amounts of ligand may have effect. Thus, dosage ranges would ordinarily be expected to be in amounts lower than 1 mM concentrations, typically less than about 10  $\mu$ M concentrations, usually less than about 100 nM, 25 preferably less than about 10 pM (picomolar), and most preferably less than about 1 fM (femtomolar), with an appropriate carrier. Slow release formulations, or slow release apparatus will often be utilized for continuous administration.

30 DTLRs, fragments thereof, and antibodies or its fragments, antagonists, and agonists, may be administered directly to the host to be treated or, depending on the size of the compounds, it may be desirable to conjugate them to carrier proteins such as ovalbumin or serum 35 albumin prior to their administration. Therapeutic formulations may be administered in any conventional dosage formulation. While it is possible for the active

ingredient to be administered alone, it is preferable to present it as a pharmaceutical formulation. Formulations comprise at least one active ingredient, as defined above, together with one or more acceptable carriers thereof. Each carrier must be both pharmaceutically and physiologically acceptable in the sense of being compatible with the other ingredients and not injurious to the patient. Formulations include those suitable for oral, rectal, nasal, or parenteral (including 5 subcutaneous, intramuscular, intravenous and intradermal) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any methods well known in the art of pharmacy. See, e.g., Gilman, et al. (eds) (1990) Goodman and Gilman's: The 10 Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences (current edition), Mack Publishing Co., Easton, Penn.; Avis, et al. (eds. 1993) Pharmaceutical Dosage Forms: Parenteral Medications Dekker, NY; Lieberman, et al. (eds. 1990) 15 Pharmaceutical Dosage Forms: Tablets Dekker, NY; and Lieberman, et al. (eds. 1990) Pharmaceutical Dosage Forms: Disperse Systems Dekker, NY. The therapy of this invention may be combined with or used in association 20 with other therapeutic agents, particularly agonists or antagonists of other IL-1 family members.

#### IX. Ligands

The description of the Toll receptors herein provide means to identify ligands, as described above. Such 30 ligand should bind specifically to the respective receptor with reasonably high affinity. Various constructs are made available which allow either labeling of the receptor to detect its ligand. For example, directly labeling DTLR, fusing onto it markers for 35 secondary labeling, e.g., FLAG or other epitope tags, etc., will allow detection of receptor. This can be histological, as an affinity method for biochemical

purification, or labeling or selection in an expression cloning approach. A two-hybrid selection system may also be applied making appropriate constructs with the available DTLR sequences. See, e.g., Fields and Song  
5 (1989) Nature 340:245-246.

Generally, descriptions of DTLRs will be analogously applicable to individual specific embodiments directed to DTLR2, DTLR3, DTLR4, DTLR5, DTLR6, DTLR7, DTLR8, DTLR9, and/or DTLR10 reagents and compositions.

10 The broad scope of this invention is best understood with reference to the following examples, which are not intended to limit the inventions to the specific embodiments.

15 EXAMPLES

I. General Methods

Some of the standard methods are described or referenced, e.g., in Maniatis, et al. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press; Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual, (2d ed.), vols 1-3, CSH Press, NY; Ausubel, et al., Biology, Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (1987 and Supplements) Current Protocols in Molecular Biology, Greene/Wiley, New York. Methods for protein purification include such methods as ammonium sulfate precipitation, column chromatography, electrophoresis, centrifugation, crystallization, and others. See, e.g., Ausubel, et al. (1987 and periodic supplements); Coligan, et al. (ed. 1996) and periodic supplements, Current Protocols In Protein Science Greene/Wiley, New York; Deutscher (1990) "Guide to Protein Purification" in Methods in Enzymology, vol. 182, and other volumes in this series; and manufacturer's literature on use of protein purification products, e.g., Pharmacia, Piscataway, N.J., or Bio-Rad, Richmond, CA.

Combination with recombinant techniques allow fusion to appropriate segments, e.g., to a FLAG sequence or an equivalent which can be fused via a protease-removable sequence. See, e.g., Hochuli (1989) Chemische Industrie

- 5 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) Genetic Engineering, Principle and Methods 12:87-98, Plenum Press, N.Y.; and Crowe, et al. (1992) OIAexpress: The High Level Expression & Protein Purification System
- 10 QUIAGEN, Inc., Chatsworth, CA.

Standard immunological techniques and assays are described, e.g., in Hertzenberg, et al. (eds. 1996) Weir's Handbook of Experimental Immunology vols. 1-4, Blackwell Science; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Methods in Enzymology volumes. 70, 73, 74, 84, 92, 93, 108, 116, 121, 132, 150, 162, and 163.

Assays for vascular biological activities are well known in the art. They will cover angiogenic and angiostatic activities in tumor, or other tissues, e.g., arterial smooth muscle proliferation (see, e.g., Koyoma, et al. (1996) Cell 87:1069-1078), monocyte adhesion to vascular epithelium (see McEvoy, et al. (1997) J. Exp. Med. 185:2069-2077), etc. See also Ross (1993) Nature 362:801-809; Rekhter and Gordon (1995) Am. J. Pathol. 147:668-677; Thyberg, et al. (1990) Atherosclerosis 10:966-990; and Gumbiner (1996) Cell 84:345-357.

Assays for neural cell biological activities are described, e.g., in Wouterlood (ed. 1995) Neuroscience Protocols modules 10, Elsevier; Methods in Neurosciences Academic Press; and Neuromethods Humana Press, Totowa, NJ. Methodology of developmental systems is described, e.g., in Meissami (ed.) Handbook of Human Growth and Developmental Biology CRC Press; and Chrispeels (ed.) Molecular Techniques and Approaches in Developmental Biology Interscience.

Computer sequence analysis is performed, e.g., using available software programs, including those from the GCG (U. Wisconsin) and GenBank sources. Public sequence databases were also used, e.g., from GenBank, NCBI, EMBO, 5 and others.

Many techniques applicable to IL-10 receptors may be applied to DTLRs, as described, e.g., in USSN 08/110,683 (IL-10 receptor), which is incorporated herein by reference for all purposes.

10

## II. Novel Family of Human Receptors

Abbreviations: DTLR, Toll-like receptor; IL-1R, interleukin-1 receptor; TH, Toll homology; LRR, leucine-rich repeat; EST, expressed sequence tag; STS, sequence tagged site; FISH, fluorescence in situ hybridization.

The discovery of sequence homology between the cytoplasmic domains of *Drosophila* Toll and human interleukin-1 (IL-1) receptors has sown the conviction that both molecules trigger related signaling pathways tied to the nuclear translocation of Rel-type transcription factors. This conserved signaling scheme governs an evolutionarily ancient immune response in both insects and vertebrates. We report the molecular cloning of a novel class of putative human receptors with a protein architecture that is closely similar to *Drosophila* Toll in both intra- and extra-cellular segments. Five human Toll-like receptors, designated 20 DTLRs 1-5, are likely the direct homologs of the fly molecule, and as such could constitute an important and unrecognition component of innate immunity in humans; intriguingly, the evolutionary retention of DTLRs in vertebrates may indicate another role, akin to Toll in 25 the dorso-ventralization of the *Drosophila* embryo, as regulators of early morphogenetic patterning. Multiple tissue mRNA blots indicate markedly different patterns of 30 35

expression for the human DTLRs. Using fluorescence *in situ* hybridization and Sequence-Tagged Site database analyses, we also show that the cognate DTLR genes reside on chromosomes 4 (DTLRs 1, 2, and 3), 9 (DTLR4), and 1 5 (DTLR5). Structure prediction of the aligned Toll-homology (TH) domains from varied insect and human DTLRs, vertebrate IL-1 receptors, and MyD88 factors, and plant disease resistance proteins, recognizes a parallel  $\beta/\alpha$  fold with an acidic active site; a similar structure 10 notably recurs in a class of response regulators broadly involved in transducing sensory information in bacteria.

The seeds of the morphogenetic gulf that so dramatically separates flies from humans are planted in 15 familiar embryonic shapes and patterns, but give rise to very different cell complexities. DeRobertis and Sasai (1996) Nature 380:37-40; and Arendt and Nübler-Jung (1997) Mech. Develop. 61:7-21. This divergence of developmental plans between insects and vertebrates is 20 choreographed by remarkably similar signaling pathways, underscoring a greater conservation of protein networks and biochemical mechanisms from unequal gene repertoires. Miklos and Rubin (1996) Cell 86:521-529; and Chothia (1994) Develop. 1994 Suppl., 27-33. A powerful way to 25 chart the evolutionary design of these regulatory pathways is by inferring their likely molecular components (and biological functions) through interspecies comparisons of protein sequences and structures. Miklos and Rubin (1996) Cell 86:521-529; 30 Chothia (1994) Develop. 1994 Suppl., 27-33 (3-5); and Banfi, et al. (1996) Nature Genet. 13:167-174.

A universally critical step in embryonic development is the specification of body axes, either born from innate asymmetries or triggered by external cues. 35 DeRobertis and Sasai (1996) Nature 380:37-40; and Arendt and Nübler-Jung (1997) Mech. Develop. 61:7-21. As a model system, particular attention has been focused on

the phylogenetic basis and cellular mechanisms of dorsoventral polarization . DeRobertis and Sasai (1996) Nature 380:37-40; and Arendt and Nübler-Jung (1997) Mech. Develop. 61:7-21. A prototype molecular strategy for 5 this transformation has emerged from the Drosophila embryo, where the sequential action of a small number of genes results in a ventralizing gradient of the transcription factor Dorsal. St. Johnston and Nüsslein-Volhard (1992) Cell 68:201-219; and Morisato and Anderson 10 (1995) Ann. Rev. Genet. 29:371-399.

This signaling pathway centers on Toll, a transmembrane receptor that transduces the binding of a maternally-secreted ventral factor, Spätzle, into the cytoplasmic engagement of Tube, an accessory molecule, 15 and the activation of Pelle, a Ser/Thr kinase that catalyzes the dissociation of Dorsal from the inhibitor Cactus and allows migration of Dorsal to ventral nuclei (Morisato and Anderson (1995) Ann. Rev. Genet. 29:371-399; and Belvin and Anderson (1996) Ann. Rev. Cell 20 Develop. Biol. 12:393-416. The Toll pathway also controls the induction of potent antimicrobial factors in the adult fly (Lemaitre, et al. (1996) Cell 86:973-983); this role in Drosophila immune defense strengthens mechanistic parallels to IL-1 pathways that govern a host 25 of immune and inflammatory responses in vertebrates. Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; and Wasserman (1993) Molec. Biol. Cell 4:767-771. A Toll-related cytoplasmic domain in IL-1 receptors directs the binding of a Pelle-like kinase, IRAK, and the 30 activation of a latent NF-κB/I-κB complex that mirrors the embrace of Dorsal and Cactus. Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; and Wasserman (1993) Molec. Biol. Cell 4:767-771.

We describe the cloning and molecular 35 characterization of four new Toll-like molecules in humans, designated DTLRs 2-5 (following Chiang & Beachy (1994) Mech. Develop. 47:225-239), that reveal a receptor

family more closely tied to *Drosophila* Toll homologs than to vertebrate IL-1 receptors. The DTLR sequences are derived from human ESTs; these partial cDNAs were used to draw complete expression profiles in human tissues for 5 the five DTLRs, map the chromosomal locations of cognate genes, and narrow the choice of cDNA libraries for full-length cDNA retrievals. Spurred by other efforts (Banfi, et al. (1996) *Nature Genet.* 13:167-174; and Wang, et al. (1996) *J. Biol. Chem.* 271:4468-4476), we are assembling, 10 by structural conservation and molecular parsimony, a biological system in humans that is the counterpart of a compelling regulatory scheme in *Drosophila*. In addition, a biochemical mechanism driving Toll signaling is suggested by the proposed tertiary fold of the Toll- 15 homology (TH) domain, a core module shared by DTLRs, a broad family of IL-1 receptors, mammalian MyD88 factors and plant disease resistance proteins. Mitcham, et al. (1996) *J. Biol. Chem.* 271:5777-5783; and Hardiman, et al. (1996) *Oncogene* 13:2467-2475. We propose that a 20 signaling route coupling morphogenesis and primitive immunity in insects, plants, and animals (Belvin and Anderson (1996) *Ann. Rev. Cell Develop. Biol.* 12:393-416; and Wilson, et al. (1997) *Curr. Biol.* 7:175-178) may have roots in bacterial two-component pathways.

25

#### Computational Analysis.

Human sequences related to insect DTLRs were identified from the EST database (dbEST) at the National Center for Biotechnology Information (NCBI) using the 30 BLAST server (Altschul, et al. (1994) *Nature Genet.* 6:119-129). More sensitive pattern- and profile-based methods (Bork and Gibson (1996) *Meth. Enzymol.* 266:162-184) were used to isolate the signaling domains of the DTLR family that are shared with vertebrate and plant 35 proteins present in nonredundant databases. The progressive alignment of DTLR intra- or extracellular domain sequences was carried out by ClustalW (Thompson,

et al. (1994) Nucleic Acids Res. 22:4673-4680); this program also calculated the branching order of aligned sequences by the Neighbor-Joining algorithm (5000 bootstrap replications provided confidence values for the tree groupings).

Conserved alignment patterns, discerned at several degrees of stringency, were drawn by the Consensus program (internet URL <http://www.bork.embl-heidelberg.de/Alignment/consensus.html>). The PRINTS library of protein fingerprints (<http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html>) (Attwood, et al. (1997) Nucleic Acids Res. 25:212-217) reliably identified the myriad leucine-rich repeats (LRRs) present in the extracellular segments of DTLRs with a compound motif (PRINTS code Leurichrpt) that flexibly matches N- and C-terminal features of divergent LRRs. Two prediction algorithms whose three-state accuracy is above 72% were used to derive a consensus secondary structure for the intracellular domain alignment, as a bridge to fold recognition efforts (Fischer, et al. (1996) FASEB J. 10:126-136). Both the neural network program PHD (Rost and Sander (1994) Proteins 19:55-72) and the statistical prediction method DSC (King and Sternberg (1996) Protein Sci. 5:2298-2310) have internet servers (URLs [http://www.embl-heidelberg.de/predictprotein/phd\\_pred.html](http://www.embl-heidelberg.de/predictprotein/phd_pred.html) and [http://bonsai.lif.icnet.uk/bmm/dsc/dsc\\_read\\_align.html](http://bonsai.lif.icnet.uk/bmm/dsc/dsc_read_align.html), respectively). The intracellular region encodes the THD region discussed, e.g., in Hardiman, et al. (1996) Oncogene 13:2467-2475; and Rock, et al. (1998) Proc. Nat'l Acad. Sci. USA 95:588-593, each of which is incorporated herein by reference. This domain is very important in the mechanism of signaling by the receptors, which transfers a phosphate group to a substrate.

35

Cloning of full-length human DTLR cDNAs.

PCR primers derived from the Toll-like Humrsc786 sequence (Genbank accession code D13637) (Nomura, et al. (1994) DNA Res 1:27-35) were used to probe a human erythroleukemic, TF-1 cell line-derived cDNA library (Kitamura, et al. (1989) Blood 73:375-380) to yield the DTLR1 cDNA sequence. The remaining DTLR sequences were flagged from dbEST, and the relevant EST clones obtained from the I.M.A.G.E. consortium (Lennon, et al. (1996) Genomics 33:151-152) via Research Genetics (Huntsville, AL): CloneID#'s 80633 and 117262 (DTLR2), 144675 (DTLR3), 202057 (DTLR4) and 277229 (DTLR5). Full length cDNAs for human DTLRs 2-4 were cloned by DNA hybridization screening of λgt10 phage, human adult lung, placenta, and fetal liver 5'-Stretch Plus cDNA libraries (Clontech), respectively; the DTLR5 sequence is derived from a human multiple-sclerosis plaque EST. All positive clones were sequenced and aligned to identify individual DTLR ORFs: DTLR1 (2366 bp clone, 786 aa ORF), DTLR2 (2600 bp, 784 aa), DTLR3 (3029 bp, 904 aa), DTLR4 (3811 bp, 879 aa) and DTLR5 (1275 bp, 370 aa). Probes for DTLR3 and DTLR4 hybridizations were generated by PCR using human placenta (Stratagene) and adult liver (Clontech) cDNA libraries as templates, respectively; primer pairs were derived from the respective EST sequences. PCR reactions were conducted using *T. aquaticus* Taqplus DNA polymerase (Stratagene) under the following conditions: 1 x (94° C, 2 min) 30 x (55° C, 20 sec; 72° C 30 sec; 94° C 20 sec), 1 x (72° C, 8 min). For DTLR2 full-length cDNA screening, a 900 bp fragment generated by EcoRI/XbaI digestion of the first EST clone (ID# 80633) was used as a probe.

mRNA blots and chromosomal localization.

Human multiple tissue (Cat# 1, 2) and cancer cell line blots (Cat# 7757-1), containing approximately 2 µg of poly(A)<sup>+</sup> RNA per lane, were purchased from Clontech (Palo Alto, CA). For DTLRs 1-4, the isolated full-length

- cDNAs served as probes, for DTLR5 the EST clone (ID #277229) plasmid insert was used. Briefly, the probes were radiolabeled with [ $\alpha$ -<sup>32</sup>P] dATP using the Amersham Rediprime random primer labeling kit (RPN1633).
- 5 Prehybridization and hybridizations were performed at 65° C in 0.5 M Na<sub>2</sub>HPO<sub>4</sub>, 7% SDS, 0.5 M EDTA (pH 8.0). All stringency washes were conducted at 65° C with two initial washes in 2 x SSC, 0.1% SDS for 40 min followed by a subsequent wash in 0.1 x SSC, 0.1% SDS for 20 min.
- 10 Membranes were then exposed at -70° C to X-Ray film (Kodak) in the presence of intensifying screens. More detailed studies by cDNA library Southern (14) were performed with selected human DTLR clones to examine their expression in hemopoietic cell subsets.
- 15 Human chromosomal mapping was conducted by the method of fluorescence in situ hybridization (FISH) as described in Heng and Tsui (1994) Meth. Molec. Biol. 33:109-122, using the various full-length (DTLRs 2-4) or partial (DTLR5) cDNA clones as probes. These analyses
- 20 were performed as a service by SeeDNA Biotech Inc. (Ontario, Canada). A search for human syndromes (or mouse defects in syntenic loci) associated with the mapped DTLR genes was conducted in the Dysmorphic Human-Mouse Homology Database by internet server
- 25 ([http://www.hgmp.mrc.ac.uk/DHMHD/hum\\_chromel.html](http://www.hgmp.mrc.ac.uk/DHMHD/hum_chromel.html)).

Conserved architecture of insect and human DTLR ectodomains.

The Toll family in *Drosophila* comprises at least 30 four distinct gene products: Toll, the prototype receptor involved in dorsoventral patterning of the fly embryo (Morisato and Anderson (1995) Ann. Rev. Genet. 29:371-399) and a second named '18 Wheeler' (18w) that may also be involved in early embryonic development (Chiang and Beachy (1994) Mech. Develop. 47:225-239; Eldon, et al. (1994) Develop. 120:885-899); two additional receptors are predicted by incomplete, Toll-like ORFs downstream of

the male-specific-transcript (Mst) locus (Genbank code X67703) or encoded by the 'sequence-tagged-site' (STS) Dm2245 (Genbank code G01378) (Mitcham, et al. (1996) J. Biol. Chem. 271:5777-5783). The extracellular segments of Toll and 18w are distinctively composed of imperfect, ~24 amino acid LRR motifs (Chiang and Beachy (1994) Mech. Develop. 47:225-239; and Eldon, et al. (1994) Develop. 120:885-899). Similar tandem arrays of LRRs commonly form the adhesive antennae of varied cell surface molecules and their generic tertiary structure is presumed to mimic the horseshoe-shaped cradle of a ribonuclease inhibitor fold, where seventeen LRRs show a repeating  $\beta/\alpha$ -hairpin, 28 residue motif (Buchanan and Gay (1996) Prog. Biophys. Molec. Biol. 65:1-44). The specific recognition of Spätzle by Toll may follow a model proposed for the binding of cystine-knot fold glycoprotein hormones by the multi-LRR ectodomains of serpentine receptors, using the concave side of the curved  $\beta$ -sheet (Kajava, et al. (1995) Structure 3:867-877); intriguingly, the pattern of cysteines in Spätzle, and an orphan Drosophila ligand, Trunk, predict a similar cystine-knot tertiary structure (Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; and Casanova, et al. (1995) Genes Develop. 9:2539-2544).

The 22 and 31 LRR ectodomains of Toll and 18w, respectively (the Mst ORF fragment displays 16 LRRs), are most closely related to the comparable 18, 19, 24, and 22 LRR arrays of DTLRs 1-4 (the incomplete DTLR5 chain presently includes four membrane-proximal LRRs) by sequence and pattern analysis (Altschul, et al. (1994) Nature Genet. 6:119-129; and Bork and Gibson (1996) Meth. Enzymol. 266:162-184) (Fig. 1). However, a striking difference in the human DTLR chains is the common loss of a ~90 residue cysteine-rich region that is variably embedded in the ectodomains of Toll, 18w and the Mst ORF (distanced four, six and two LRRs, respectively, from the membrane boundary). These cysteine clusters are

bipartite, with distinct 'top' (ending an LRR) and 'bottom' (stacked atop an LRR) halves (Chiang and Beachy (1994) Mech. Develop. 47:225-239; Eldon, et al. (1994) Develop. 120:885-899; and Buchanan and Gay (1996) Prog. Biophys. Molec. Biol. 65:1-44); the 'top' module recurs in both Drosophila and human DTLRs as a conserved juxtamembrane spacer (Fig. 1). We suggest that the flexibly located cysteine clusters in Drosophila receptors (and other LRR proteins), when mated 'top' to 'bottom', form a compact module with paired termini that can be inserted between any pair of LRRs without altering the overall fold of DTLR ectodomains; analogous 'extruded' domains decorate the structures of other proteins (Russell (1994) Protein Engin. 7:1407-1410).

15

Molecular design of the TH signaling domain.

Sequence comparison of Toll and IL-1 type-I (IL-1R1) receptors has disclosed a distant resemblance of a ~200 amino acid cytoplasmic domain that presumably mediates 20 signaling by similar Rel-type transcription factors. Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; and (Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; and Wasserman (1993) Molec. Biol. Cell 4:767-771). More recent additions to 25 this functional paradigm include a pair of plant disease resistance proteins from tobacco and flax that feature an N-terminal TH module followed by nucleotide-binding (NTPase) and LRR segments (Wilson, et al. (1997) Curr. Biol. 7:175-178); by contrast, a 'death domain' precedes 30 the TH chain of MyD88, an intracellular myeloid differentiation marker (Mitcham, et al. (1996) J. Biol. Chem. 271:5777-5783; and Hardiman, et al. (1996) Oncogene 13:2467-2475) (Fig. 1). New IL-1-type receptors include 35 IL-1R3, an accessory signaling molecule, and orphan receptors IL-1R4 (also called ST2/Fit-1/T1), IL-1R5 (IL-1R-related protein), and IL-1R6 (IL-1R-related protein-2) (Mitcham, et al. (1996) J. Biol. Chem. 271:5777-

5783;Hardiman, et al. (1996) Oncogene 13:2467-2475). With the new human DTLR sequences, we have sought a structural definition of this evolutionary thread by analyzing the conformation of the common TH module: ten 5 blocks of conserved sequence comprising 128 amino acids form the minimal TH domain fold; gaps in the alignment mark the likely location of sequence and length-variable loops (Fig. 2a).

Two prediction algorithms that take advantage of the 10 patterns of conservation and variation in multiply aligned sequences, PHD (Rost and Sander (1994) Proteins 19:55-72) and DSC (King and Sternberg (1996) Protein Sci. 5:2298-2310), produced strong, concordant results for the TH signaling module (Fig. 2a). Each block contains a 15 discrete secondary structural element: the imprint of alternating  $\beta$ -strands (labeled A-E) and  $\alpha$ -helices (numbered 1-5) is diagnostic of an  $\beta/\alpha$ -class fold with  $\alpha$ -helices on both faces of a parallel  $\beta$ -sheet. Hydrophobic  $\beta$ -strands A, C and D are predicted to form 'interior' 20 staves in the  $\beta$ -sheet, while the shorter, amphipathic  $\beta$ -strands B and E resemble typical 'edge' units (Fig. 2a). This assignment is consistent with a strand order of B-A-C-D-E in the core  $\beta$ -sheet (Fig. 2b); fold comparison 25 ('mapping') and recognition ('threading') programs (Fischer, et al. (1996) FASEB J. 10:126-136) strongly return this doubly wound  $\beta/\alpha$  topology. A surprising, functional prediction of this outline structure for the 30 TH domain is that many of the conserved, charged residues in the multiple alignment map to the C-terminal end of the  $\beta$ -sheet: residue Asp16 (block numbering scheme - Fig. 2a) at the end of  $\beta$ A, Arg39 and Asp40 following  $\beta$ B, Glu75 in the first turn of  $\alpha$ 3, and the more loosely conserved Glu/Asp residues in the  $\beta$ D- $\alpha$ 4 loop, or after  $\beta$ E (Fig. 2a). The location of four other conserved residues 35 (Asp7, Glu28, and the Arg57-Arg/Lys58 pair) is compatible with a salt bridge network at the opposite, N-terminal end of the  $\beta$ -sheet (Fig. 2a).

Signaling function depends on the structural integrity of the TH domain. Inactivating mutations or deletions within the module boundaries (Fig. 2a) have been catalogued for IL-1R1 and Toll. Heguy, et al.

5 (1992) J. Biol. Chem. 267:2605-2609; Croston, et al.  
(1995) J. Biol. Chem. 270:16514-16517; Schneider, et al.  
(1991) Genes Develop. 5:797-807; Norris and Manley.  
(1992) Genes Develop. 6:1654-1667; Norris and Manley  
(1995) Genes Develop. 9:358-369; and Norris and Manley

10 (1996) Genes Develop. 10:862-872. The human DTLR1-5 chains extending past the minimal TH domain (8, 0, 6, 22 and 18 residue lengths, respectively) are most closely similar to the stubby, 4 aa 'tail' of the Mst ORF. Toll and 18w display unrelated 102 and 207 residue tails (Fig.

15 2a) that may negatively regulate the signaling of the fused TH domains. Norris and Manley (1995) Genes Develop. 9:358-369; and Norris and Manley (1996) Genes Develop. 10:862-872.

The evolutionary relationship between the disparate proteins that carry the TH domain can best be discerned by a phylogenetic tree derived from the multiple alignment (Fig. 3). Four principal branches segregate the plant proteins, the MyD88 factors, IL-1 receptors and Toll-like molecules; the latter branch clusters the

20 Drosophila and human DTLRs.

#### Chromosomal dispersal of human DTLR genes.

In order to investigate the genetic linkage of the nascent human DTLR gene family, we mapped the chromosomal loci of four of the five genes by FISH (Fig. 4). The DTLR1 gene has previously been charted by the human genome project: an STS database locus (dbSTS accession number G06709, corresponding to STS WI-7804 or SHGC-12827) exists for the Humrsc786 cDNA (Nomura, et al.

30 (1994) DNA Res 1:27-35) and fixes the gene to chromosome 4 marker interval D4S1587-D42405 (50-56 cM) circa 4p14.

35 This assignment has recently been corroborated by FISH

analysis. Taguchi, et al. (1996) Genomics 32:486-488. In the present work, we reliably assign the remaining DTLR genes to loci on chromosome 4q32 (DTLR2), 4q35 (DTLR3), 9q32-33 (DTLR4) and 1q33.3 (DTLR5). During the 5 course of this work, an STS for the parent DTLR2 EST (cloneID # 80633) has been generated (dbSTS accession number T57791 for STS SHGC-33147) and maps to the chromosome 4 marker interval D4S424-D4S1548 (143-153 cM) at 4q32 -in accord with our findings. There is a ~50 cM 10 gap between DTLR2 and DTLR3 genes on the long arm of chromosome 4.

DTLR genes are differentially expressed.

Both Toll and 18w have complex spatial and temporal 15 patterns of expression in *Drosophila* that may point to functions beyond embryonic patterning. St. Johnston and Nüsslein-Volhard (1992) Cell 68:201-219; Morisato and Anderson (1995) Ann. Rev. Genet. 29:371-399; Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; 20 Lemaitre, et al. (1996) Cell 86:973-983; Chiang and Beachy (1994) Mech. Develop. 47:225-239; and Eldon, et al. (1994) Develop. 120:885-899. We have examined the spatial distribution of DTLR transcripts by mRNA blot analysis with varied human tissue and cancer cell lines 25 using radioabeled DTLR cDNAs (Fig. 5). DTLR1 is found to be ubiquitously expressed, and at higher levels than the other receptors. Presumably reflecting alternative splicing, 'short' 3.0 kB and 'long' 8.0 kB DTLR1 transcript forms are present in ovary and spleen, respectively (Fig. 5, panels A & B). A cancer cell mRNA 30 panel also shows the prominent overexpression of DTLR1 in a Burkitt's Lymphoma Raji cell line (Fig. 5, panel C). DTLR2 mRNA is less widely expressed than DTLR1, with a 4.0 kB species detected in lung and a 4.4 kB transcript 35 evident in heart, brain and muscle. The tissue distribution pattern of DTLR3 echoes that of DTLR2 (Fig. 5, panel E). DTLR3 is also present as two major

transcripts of approximately 4.0 and 6.0 kB in size, and the highest levels of expression are observed in placenta and pancreas. By contrast, DTLR4 and DTLR5 messages appear to be extremely tissue-specific. DTLR4 was 5 detected only in placenta as a single transcript of ~7.0 kB in size. A faint 4.0 kB signal was observed for DTLR5 in ovary and peripheral blood monocytes.

Components of an evolutionarily ancient regulatory 10 system.

The original molecular blueprints and divergent fates of signaling pathways can be reconstructed by comparative genomic approaches. Miklos and Rubin (1996) Cell 86:521-529; Chothia (1994) Develop. 1994 Suppl., 27-15 33; Banfi, et al. (1996) Nature Genet. 13:167-174; and Wang, et al. (1996) J. Biol. Chem. 271:4468-4476. We have used this logic to identify an emergent gene family 20 in humans, encoding five receptor paralogs at present, DTLRs 1-5, that are the direct evolutionary counterparts of a *Drosophila* gene family headed by Toll (Figs. 1-3). The conserved architecture of human and fly DTLRs, conserved LRR ectodomains and intracellular TH modules (Fig. 1), intimates that the robust pathway coupled to Toll in *Drosophila* (6, 7) survives in vertebrates. The 25 best evidence borrows from a reiterated pathway: the manifold IL-1 system and its repertoire of receptor-fused TH domains, IRAK, NF- $\kappa$ B and I- $\kappa$ B homologs (Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; Wasserman (1993) Molec. Biol. Cell 4:767-771; Hardiman, 30 et al. (1996) Oncogene 13:2467-2475; and Cao, et al. (1996) Science 271:1128-1131); a Tube-like factor has also been characterized. It is not known whether DTLRs can productively couple to the IL-1R signaling machinery, or instead, a parallel set of proteins is used. 35 Differently from IL-1 receptors, the LRR cradle of human DTLRs is predicted to retain an affinity for Spätzle/Trunk-related cystine-knot factors; candidate

DTLR ligands (called PENS) that fit this mold have been isolated.

Biochemical mechanisms of signal transduction can be gauged by the conservation of interacting protein folds in a pathway. Miklos and Rubin (1996) Cell 86:521-529; Chothia (1994) Develop. 1994 Suppl., 27-33. At present, the Toll signaling paradigm involves some molecules whose roles are narrowly defined by their structures, actions or fates: Pelle is a Ser/Thr kinase (phosphorylation), Dorsal is an NF- $\kappa$ B-like transcription factor (DNA-binding) and Cactus is an ankyrin-repeat inhibitor (Dorsal binding, degradation). Belvin and Anderson (1996) Ann. Rev. Cell Develop. Biol. 12:393-416. By contrast, the functions of the Toll TH domain and Tube remain enigmatic. Like other cytokine receptors (Heldin (1995) Cell 80:213-223), ligand-mediated dimerization of Toll appears to be the triggering event: free cysteines in the juxtamembrane region of Toll create constitutively active receptor pairs (Schneider, et al. (1991) Genes Develop. 5:797-807), and chimeric Torso-Toll receptors signal as dimers (Galindo, et al. (1995) Develop. 121:2209-2218); yet, severe truncations or wholesale loss of the Toll ectodomain results in promiscuous intracellular signaling (Norris and Manley (1995) Genes Develop. 9:358-369; and Winans and Hashimoto (1995) Molec. Biol. Cell 6:587-596), reminiscent of oncogenic receptors with catalytic domains (Heldin (1995) Cell 80:213-223). Tube is membrane-localized, engages the N-terminal (death) domain of Pelle and is phosphorylated, but neither Toll-Tube or Toll-Pelle interactions are registered by two-hybrid analysis (Galindo, et al. (1995) Develop. 121:2209-2218; and Großhans, et al. (1994) Nature 372:563-566); this latter result suggests that the conformational 'state' of the Toll TH domain somehow affects factor recruitment. Norris and Manley (1996) Genes Develop. 10:862-872; and Galindo, et al. (1995) Develop. 121:2209-2218.

At the heart of these vexing issues is the structural nature of the Toll TH module. To address this question, we have taken advantage of the evolutionary diversity of TH sequences from insects, plants and vertebrates, incorporating the human DTLR chains, and extracted the minimal, conserved protein core for structure prediction and fold recognition (Fig. 2). The strongly predicted ( $\beta/\alpha$ )<sub>5</sub> TH domain fold with its asymmetric cluster of acidic residues is topologically identical to the structures of response regulators in bacterial two-component signaling pathways (Volz (1993) *Biochemistry* 32:11741-11753; and Parkinson (1993) *Cell* 73:857-871) (Fig. 2). The prototype chemotaxis regulator CheY transiently binds a divalent cation in an 'aspartate pocket' at the C-end of the core  $\beta$ -sheet; this cation provides electrostatic stability and facilitates the activating phosphorylation of an invariant Asp. Volz (1993) *Biochemistry* 32:11741-11753. Likewise, the TH domain may capture cations in its acidic nest, but activation, and downstream signaling, could depend on the specific binding of a negatively charged moiety: anionic ligands can overcome intensely negative binding-site potentials by locking into precise hydrogen-bond networks. Ledvina, et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:6786-6791. Intriguingly, the TH domain may not simply act as a passive scaffold for the assembly of a Tube/Pelle complex for Toll, or homologous systems in plants and vertebrates, but instead actively participate as a true conformational trigger in the signal transducing machinery. Perhaps explaining the conditional binding of a Tube/Pelle complex, Toll dimerization could promote unmasking, by regulatory receptor tails (Norris and Manley (1995) *Genes Develop.* 9:358-369; Norris and Manley (1996) *Genes Develop.* 10:862-872), or binding by small molecule activators of the TH pocket. However, 'free' TH modules inside the cell (Norris and Manley (1995) *Genes Develop.* 9:358-369;

Winans and Hashimoto (1995) Molec. Biol. 6:587-596) could act as catalytic, CheY-like triggers by activating and docking with errant Tube/Pelle complexes.

5 Morphogenetic receptors and immune defense.

The evolutionary link between insect and vertebrate immune systems is stamped in DNA: genes encoding antimicrobial factors in insects display upstream motifs similar to acute phase response elements known to bind 10 NF- $\kappa$ B transcription factors in mammals. Hultmark (1993) Trends Genet. 9:178-183. Dorsal, and two Dorsal-related factors, Dif and Relish, help induce these defense proteins after bacterial challenge (Reichhart, et al. (1993) C. R. Acad. Sci. Paris 316:1218-1224; Ip, et al. 15 (1993) Cell 75:753-763; and Dushay, et al. (1996) Proc. Natl. Acad. Sci. USA 93:10343-10347); Toll, or other DTLRs, likely modulate these rapid immune responses in adult Drosophila (Lemaitre, et al. (1996) Cell 86:973-983; and Rosetto, et al. (1995) Biochem. Biophys. Res. Commun. 209:111-116). These mechanistic parallels to the IL-1 inflammatory response in vertebrates are evidence of the functional versatility of the Toll signaling pathway, and suggest an ancient synergy between embryonic patterning and innate immunity (Belvin and Anderson 20 (1996) Ann. Rev. Cell Develop. Biol. 12:393-416; Lemaitre, et al. (1996) Cell 86:973-983; Wasserman (1993) Molec. Biol. Cell 4:767-771; Wilson, et al. (1997) Curr. Biol. 7:175-178; Hultmark (1993) Trends Genet. 9:178-183; Reichhart, et al. (1993) C. R. Acad. Sci. Paris 316:1218- 25 1224; Ip, et al. (1993) Cell 75:753-763; Dushay, et al. (1996) Proc. Natl. Acad. Sci. USA 93:10343-10347; Rosetto, et al. (1995) Biochem. Biophys. Res. Commun. 209:111-116; Medzhitov and Janeway (1997) Curr. Opin. Immunol. 9:4-9; and Medzhitov and Janeway (1997) Curr. Opin. Immunol. 9:4-9). The closer homology of insect and 30 human DTLR proteins invites an even stronger overlap of biological functions that supersedes the purely immune

parallels to IL-1 systems, and lends potential molecular regulators to dorso-ventral and other transformations of vertebrate embryos. DeRobertis and Sasai (1996) Nature 380:37-40; and Arendt and Nübler-Jung (1997) Mech. 5 Develop. 61:7-21.

The present description of an emergent, robust receptor family in humans mirrors the recent discovery of the vertebrate Frizzled receptors for Wnt patterning factors. Wang, et al. (1996) J. Biol. Chem. 271:4468-10 4476. As numerous other cytokine-receptor systems have roles in early development (Lemaire and Kodjabachian (1996) Trends Genet. 12:525-531), perhaps the distinct cellular contexts of compact embryos and gangly adults simply result in familiar signaling pathways and their 15 diffusible triggers having different biological outcomes at different times, e.g., morphogenesis versus immune defense for DTLRs. For insect, plant, and human Toll-related systems (Hardiman, et al. (1996) Oncogene 13:2467-2475; Wilson, et al. (1997) Curr. Biol. 7:175-20 178), these signals course through a regulatory TH domain that intriguingly resembles a bacterial transducing engine (Parkinson (1993) Cell 73:857-871).

In particular, the DTLR6 exhibits structural features which establish its membership in the family. 25 Moreover, members of the family have been implicated in a number of significant developmental disease conditions and with function of the innate immune system. In particular, the DTLR6 has been mapped to the X chromosome to a location which is a hot spot for major developmental 30 abnormalities. See, e.g., The Sanger Center: human X chromosome website  
<http://www.sanger.ac.uk/HGP/ChrX/index.shtml>; and the Baylor College of Medicine Human Genome Sequencing website <http://gc.bcm.tmc.edu:8088/cgi-bin/seq/home>.

35 The accession number for the deposited PAC is AC003046. This accession number contains sequence from two PACs: RPC-164K3 and RPC-263P4. These two PAC

sequences mapped on human chromosome Xp22 at the Baylor web site between STS markers DXS704 and DXS7166. This region is a "hot spot" for severe developmental abnormalities.

5

### III. Amplification of DTLR fragment by PCR

Two appropriate primer sequences are selected (see Tables 1 through 10). RT-PCR is used on an appropriate mRNA sample selected for the presence of message to produce a partial or full length cDNA, e.g., a sample which expresses the gene. See, e.g., Innis, et al. (eds. 10 1990) PCR Protocols: A Guide to Methods and Applications Academic Press, San Diego, CA; and Dieffenbach and Dveksler (1995; eds.) PCR Primer: A Laboratory Manual 15 Cold Spring Harbor Press, CSH, NY. Such will allow determination of a useful sequence to probe for a full length gene in a cDNA library. The TLR6 is a contiguous sequence in the genome, which may suggest that the other TLRs are also. Thus, PCR on genomic DNA may yield full 20 length contiguous sequence, and chromosome walking methodology would then be applicable. Alternatively, sequence databases will contain sequence corresponding to portions of the described embodiments, or closely related forms, e.g., alternative splicing, etc. Expression 25 cloning techniques also may be applied on cDNA libraries.

### IV. Tissue distribution of DTLRs

Message for each gene encoding these DTLRs has been detected. See Figures 5A-5F. Other cells and tissues 30 will be assayed by appropriate technology, e.g., PCR, immunoassay, hybridization, or otherwise. Tissue and organ cDNA preparations are available, e.g., from Clontech, Mountain View, CA. Identification of sources of natural expression are useful, as described.

35 Southern Analysis: DNA (5 µg) from a primary amplified cDNA library is digested with appropriate restriction enzymes to release the inserts, run on a 1% agarose gel and

transferred to a nylon membrane (Schleicher and Schuell, Keene, NH).

Samples for human mRNA isolation would typically include, e.g.: peripheral blood mononuclear cells (monocytes, T cells, NK cells, granulocytes, B cells), resting (T100); peripheral blood mononuclear cells, activated with anti-CD3 for 2, 6, 12 h pooled (T101); T cell, TH0 clone Mot 72, resting (T102); T cell, TH0 clone Mot 72, activated with anti-CD28 and anti-CD3 for 3, 6, 12 h pooled (T103); T cell, TH0 clone Mot 72, anergic treated with specific peptide for 2, 7, 12 h pooled (T104); T cell, TH1 clone HY06, resting (T107); T cell, TH1 clone HY06, activated with anti-CD28 and anti-CD3 for 3, 6, 12 h pooled (T108); T cell, TH1 clone HY06, anergic treated with specific peptide for 2, 6, 12 h pooled (T109); T cell, TH2 clone HY935, resting (T110); T cell, TH2 clone HY935, activated with anti-CD28 and anti-CD3 for 2, 7, 12 h pooled (T111); T cells CD4+CD45RO- T cells polarized 27 days in anti-CD28, IL-4, and anti IFN- $\gamma$ , TH2 polarized, activated with anti-CD3 and anti-CD28 4 h (T116); T cell tumor lines Jurkat and Hut78, resting (T117); T cell clones, pooled AD130.2, Tc783.12, Tc783.13, Tc783.58, Tc782.69, resting (T118); T cell random  $\gamma\delta$  T cell clones, resting (T119); Splenocytes, resting (B100); Splenocytes, activated with anti-CD40 and IL-4 (B101); B cell EBV lines pooled WT49, RSB, JY, CVIR, 721.221, RM3, HSY, resting (B102); B cell line JY, activated with PMA and ionomycin for 1, 6 h pooled (B103); NK 20 clones pooled, resting (K100); NK 20 clones pooled, activated with PMA and ionomycin for 6 h (K101); NKL clone, derived from peripheral blood of LGL leukemia patient, IL-2 treated (K106); NK cytotoxic clone 640-A30-1, resting (K107); hematopoietic precursor line TF1, activated with PMA and ionomycin for 1, 6 h pooled (C100); U937 premonocytic line, resting (M100); U937 premonocytic line, activated with PMA and ionomycin for 1, 6 h pooled (M101); elutriated monocytes, activated

with LPS, IFN $\gamma$ , anti-IL-10 for 1, 2, 6, 12, 24 h pooled (M102); elutriated monocytes, activated with LPS, IFN $\gamma$ , IL-10 for 1, 2, 6, 12, 24 h pooled (M103); elutriated monocytes, activated with LPS, IFN $\gamma$ , anti-IL-10 for 4, 16  
5 h pooled (M106); elutriated monocytes, activated with LPS, IFN $\gamma$ , IL-10 for 4, 16 h pooled (M107); elutriated monocytes, activated LPS for 1 h (M108); elutriated monocytes, activated LPS for 6 h (M109); DC 70% CD1a+, from CD34+ GM-CSF, TNF $\alpha$  12 days, resting (D101); DC 70%  
10 CD1a+, from CD34+ GM-CSF, TNF $\alpha$  12 days, activated with PMA and ionomycin for 1 hr (D102); DC 70% CD1a+, from CD34+ GM-CSF, TNF $\alpha$  12 days, activated with PMA and ionomycin for 6 hr (D103); DC 95% CD1a+, from CD34+ GM-CSF, TNF $\alpha$  12 days FACS sorted, activated with PMA and  
15 ionomycin for 1, 6 h pooled (D104); DC 95% CD14+, ex CD34+ GM-CSF, TNF $\alpha$  12 days FACS sorted, activated with PMA and ionomycin 1, 6 hr pooled (D105); DC CD1a+ CD86+, from CD34+ GM-CSF, TNF $\alpha$  12 days FACS sorted, activated with PMA and ionomycin for 1, 6 h pooled (D106); DC from  
20 monocytes GM-CSF, IL-4 5 days, resting (D107); DC from monocytes GM-CSF, IL-4 5 days, resting (D108); DC from monocytes GM-CSF, IL-4 5 days, activated LPS 4, 16 h pooled (D109); DC from monocytes GM-CSF, IL-4 5 days, activated TNF $\alpha$ , monocyte supe for 4, 16 h pooled (D110);  
25 leiomyoma L11 benign tumor (X101); normal myometrium M5 (O115); malignant leiomyosarcoma GS1 (X103); lung fibroblast sarcoma line MRC5, activated with PMA and ionomycin for 1, 6 h pooled (C101); kidney epithelial carcinoma cell line CHA, activated with PMA and ionomycin  
30 for 1, 6 h pooled (C102); kidney fetal 28 wk male (O100); lung fetal 28 wk male (O101); liver fetal 28 wk male (O102); heart fetal 28 wk male (O103); brain fetal 28 wk male (O104); gallbladder fetal 28 wk male (O106); small intestine fetal 28 wk male (O107); adipose tissue fetal  
35 28 wk male (O108); ovary fetal 25 wk female (O109); uterus fetal 25 wk female (O110); testes fetal 28 wk male

(O111); spleen fetal 28 wk male (O112); adult placenta 28 wk (O113); and tonsil inflamed, from 12 year old (X100).

Samples for mouse mRNA isolation can include, e.g.: resting mouse fibroblastic L cell line (C200); Braf:ER 5 (Braf fusion to estrogen receptor) transfected cells, control (C201); T cells, TH1 polarized (Mell14 bright, CD4+ cells from spleen, polarized for 7 days with IFN- $\gamma$  and anti IL-4; T200); T cells, TH2 polarized (Mell14 bright, CD4+ cells from spleen, polarized for 7 days with 10 IL-4 and anti-IFN- $\gamma$ ; T201); T cells, highly TH1 polarized (see Openshaw, et al. (1995) *J. Exp. Med.* 182:1357-1367; activated with anti-CD3 for 2, 6, 16 h pooled; T202); T cells, highly TH2 polarized (see Openshaw, et al. (1995) *J. Exp. Med.* 182:1357-1367; activated with anti-CD3 for 15 2, 6, 16 h pooled; T203); CD44- CD25+ pre T cells, sorted from thymus (T204); TH1 T cell clone D1.1, resting for 3 weeks after last stimulation with antigen (T205); TH1 T cell clone D1.1, 10  $\mu$ g/ml ConA stimulated 15 h (T206); TH2 T cell clone CDC35, resting for 3 weeks after last 20 stimulation with antigen (T207); TH2 T cell clone CDC35, 10  $\mu$ g/ml ConA stimulated 15 h (T208); Mell14+ naive T cells from spleen, resting (T209); Mell14+ T cells, polarized to Th1 with IFN- $\gamma$ /IL-12/anti-IL-4 for 6, 12, 24 h pooled (T210); Mell14+ T cells, polarized to Th2 with 25 IL-4/anti-IFN- $\gamma$  for 6, 13, 24 h pooled (T211); unstimulated mature B cell leukemia cell line A20 (B200); unstimulated B cell line CH12 (B201); unstimulated large B cells from spleen (B202); B cells from total spleen, LPS activated (B203); metrizamide enriched dendritic 30 cells from spleen, resting (D200); dendritic cells from bone marrow, resting (D201); monocyte cell line RAW 264.7 activated with LPS 4 h (M200); bone-marrow macrophages derived with GM and M-CSF (M201); macrophage cell line J774, resting (M202); macrophage cell line J774 + LPS + 35 anti-IL-10 at 0.5, 1, 3, 6, 12 h pooled (M203); macrophage cell line J774 + LPS + IL-10 at 0.5, 1, 3, 5, 12 h pooled(M204); aerosol challenged mouse lung tissue,

Th2 primers, aerosol OVA challenge 7, 14, 23 h pooled (see Garlisi, et al. (1995) Clinical Immunology and Immunopathology 75:75-83; X206); Nippostrongulus-infected lung tissue (see Coffman, et al. (1989) Science 245:308-310; X200); total adult lung, normal (O200); total lung, rag-1 (see Schwarz, et al. (1993) Immunodeficiency 4:249-252; O205); IL-10 K.O. spleen (see Kuhn, et al. (1991) Cell 75:263-274; X201); total adult spleen, normal (O201); total spleen, rag-1 (O207); IL-10 K.O. Peyer's patches (O202); total Peyer's patches, normal (O210); IL-10 K.O. mesenteric lymph nodes (X203); total mesenteric lymph nodes, normal (O211); IL-10 K.O. colon (X203); total colon, normal (O212); NOD mouse pancreas (see Makino, et al. (1980) Jikken Dobutsu 29:1-13; X205); 15 total thymus, rag-1 (O208); total kidney, rag-1 (O209); total heart, rag-1 (O202); total brain, rag-1 (O203); total testes, rag-1 (O204); total liver, rag-1 (O206); rat normal joint tissue (O300); and rat arthritic joint tissue (X300).

20

#### V. Cloning of species counterparts of DTLRs

Various strategies are used to obtain species counterparts of these DTLRs, preferably from other 25 primates. One method is by cross hybridization using closely related species DNA probes. It may be useful to go into evolutionarily similar species as intermediate steps. Another method is by using specific PCR primers based on the identification of blocks of similarity or 30 difference between particular species, e.g., human, genes, e.g., areas of highly conserved or nonconserved polypeptide or nucleotide sequence. Alternatively, antibodies may be used for expression cloning.

#### 35 VI. Production of mammalian DTLR protein

An appropriate, e.g., GST, fusion construct is engineered for expression, e.g., in E. coli. For

example, a mouse IGIF pGex plasmid is constructed and transformed into *E. coli*. Freshly transformed cells are grown in LB medium containing 50 µg/ml ampicillin and induced with IPTG (Sigma, St. Louis, MO). After 5 overnight induction, the bacteria are harvested and the pellets containing the DTLR protein are isolated. The pellets are homogenized in TE buffer (50 mM Tris-base pH 8.0, 10 mM EDTA and 2 mM pefabloc) in 2 liters. This material is passed through a microfluidizer 10 (Microfluidics, Newton, MA) three times. The fluidized supernatant is spun down on a Sorvall GS-3 rotor for 1 h at 13,000 rpm. The resulting supernatant containing the DTLR protein is filtered and passed over a glutathione-SEPHAROSE column equilibrated in 50 mM Tris-base pH 8.0. 15 The fractions containing the DTLR-GST fusion protein are pooled and cleaved with thrombin (Enzyme Research Laboratories, Inc., South Bend, IN). The cleaved pool is then passed over a Q-SEPHAROSE column equilibrated in 50 mM Tris-base. Fractions containing DTLR are pooled and 20 diluted in cold distilled H<sub>2</sub>O, to lower the conductivity, and passed back over a fresh Q-Sepharose column, alone or in succession with an immunoaffinity antibody column.. Fractions containing the DTLR protein are pooled, aliquoted, and stored in the -70° C freezer.

25 Comparision of the CD spectrum with DTLR1 protein may suggest that the protein is correctly folded. See Hazuda, et al. (1969) J. Biol. Chem. 264:1689-1693.

## VII. Biological Assays with DTLRs

30 Biological assays will generally be directed to the ligand binding feature of the protein or to the kinase/phosphatase activity of the receptor. The activity will typically be reversible, as are many other enzyme actions.mediate phosphatase or phosphorylase 35 activities, which activities are easily measured by standard procedures. See, e.g., Hardie, et al. (eds. 1995) The Protein Kinase FactBook vols. I and II,

Academic Press, San Diego, CA; Hanks, et al. (1991) Meth. Enzymol. 200:38-62; Hunter, et al. (1992) Cell 70:375-388; Lewin (1990) Cell 61:743-752; Pines, et al. (1991) Cold Spring Harbor Symp. Quant. Biol. 56:449-463; and  
5 Parker, et al. (1993) Nature 363:736-738.

The family of interleukins 1 contains molecules, each of which is an important mediator of inflammatory disease. For a comprehensive review, see Dinarello (1996) "Biologic basis for interleukin-1 in disease" 10 Blood 87:2095-2147. There are suggestions that the various Toll ligands may play important roles in the initiation of disease, particularly inflammatory responses. The finding of novel proteins related to the IL-1 family furthers the identification of molecules that 15 provide the molecular basis for initiation of disease and allow for the development of therapeutic strategies of increased range and efficacy.

VIII. Preparation of antibodies specific for, e.g.,  
20 DTLR4

Inbred Balb/c mice are immunized intraperitoneally with recombinant forms of the protein, e.g., purified DTLR4 or stable transfected NIH-3T3 cells. Animals are boosted at appropriate time points with protein, with or 25 without additional adjuvant, to further stimulate antibody production. Serum is collected, or hybridomas produced with harvested spleens.

Alternatively, Balb/c mice are immunized with cells transformed with the gene or fragments thereof, either 30 endogenous or exogenous cells, or with isolated membranes enriched for expression of the antigen. Serum is collected at the appropriate time, typically after numerous further administrations. Various gene therapy techniques may be useful, e.g., in producing protein in situ, for generating an immune response.

Monoclonal antibodies may be made. For example, splenocytes are fused with an appropriate fusion partner

and hybridomas are selected in growth medium by standard procedures. Hybridoma supernatants are screened for the presence of antibodies which bind to the desired DTLR, e.g., by ELISA or other assay. Antibodies which 5 specifically recognize specific DTLR embodiments may also be selected or prepared.

In another method, synthetic peptides or purified protein are presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g., Coligan 10 (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press. In appropriate situations, the binding reagent is either labeled as described above, e.g., fluorescence or otherwise, or immobilized to a 15 substrate for panning methods. Nucleic acids may also be introduced into cells in an animal to produce the antigen, which serves to elicit an immune response. See, e.g., Wang, et al. (1993) Proc. Nat'l. Acad. Sci. 90:4156-4160; Barry, et al. (1994) BioTechniques 16:616-20 619; and Xiang, et al. (1995) Immunity 2: 129-135.

#### IX. Production of fusion proteins with, e.g., DTLR5

Various fusion constructs are made with DTLR5. This portion of the gene is fused to an epitope tag, e.g., a 25 FLAG tag, or to a two hybrid system construct. See, e.g., Fields and Song (1989) Nature 340:245-246.

The epitope tag may be used in an expression cloning procedure with detection with anti-FLAG antibodies to detect a binding partner, e.g., ligand for the respective 30 DTLR5. The two hybrid system may also be used to isolate proteins which specifically bind to DTLR5.

#### X. Chromosomal mapping of DTLRs

Chromosome spreads are prepared. In situ 35 hybridization is performed on chromosome preparations obtained from phytohemagglutinin-stimulated lymphocytes cultured for 72 h. 5-bromodeoxyuridine is added for the

final seven hours of culture (60 µg/ml of medium), to ensure a posthybridization chromosomal banding of good quality.

An appropriate fragment, e.g., a PCR fragment,  
5 amplified with the help of primers on total B cell cDNA template, is cloned into an appropriate vector. The vector is labeled by nick-translation with  $^3\text{H}$ . The radiolabeled probe is hybridized to metaphase spreads as described in Mattei, et al. (1985) Hum. Genet. 69:327-  
10 331.

After coating with nuclear track emulsion (KODAK NTB2), slides are exposed, e.g., for 18 days at 4° C. To avoid any slipping of silver grains during the banding procedure, chromosome spreads are first stained with  
15 buffered Giemsa solution and metaphase photographed. R-banding is then performed by the fluorochrome-photolysis-Giemsa (FPG) method and metaphases rephotographed before analysis.

Alternatively, FISH can be performed, as described  
20 above. The DTLR genes are located on different chromosomes. DTLR2 and DTLR3 are localized to human chromosome 4; DTLR4 is localized to human chromosome 9, and DTLR5 is localized to human chromosome 1. See Figures 4A-4D.

25

#### XI. Structure activity relationship

Information on the criticality of particular residues is determined using standard procedures and analysis. Standard mutagenesis analysis is performed,  
30 e.g., by generating many different variants at determined positions, e.g., at the positions identified above, and evaluating biological activities of the variants. This may be performed to the extent of determining positions which modify activity, or to focus on specific positions  
35 to determine the residues which can be substituted to either retain, block, or modulate biological activity.

Alternatively, analysis of natural variants can indicate what positions tolerate natural mutations. This may result from populational analysis of variation among individuals, or across strains or species. Samples from 5 selected individuals are analysed, e.g., by PCR analysis and sequencing. This allows evaluation of population polymorphisms.

#### XI. Isolation of a ligand for a DTLR

10 A DTLR can be used as a specific binding reagent to identify its binding partner, by taking advantage of its specificity of binding, much like an antibody would be used. A binding reagent is either labeled as described above, e.g., fluorescence or otherwise, or immobilized to 15 a substrate for panning methods.

The binding composition is used to screen an expression library made from a cell line which expresses a binding partner, i.e., ligand, preferably membrane associated. Standard staining techniques are used to 20 detect or sort surface expressed ligand, or surface expressing transformed cells are screened by panning. Screening of intracellular expression is performed by various staining or immunofluorescence procedures. See also McMahan, et al. (1991) EMBO J. 10:2821-2832.

25 For example, on day 0, precoat 2-chamber permanox slides with 1 ml per chamber of fibronectin, 10 ng/ml in PBS, for 30 min at room temperature. Rinse once with PBS. Then plate COS cells at  $2-3 \times 10^5$  cells per chamber in 1.5 ml of growth media. Incubate overnight at 37° C.

30 On day 1 for each sample, prepare 0.5 ml of a solution of 66 µg/ml DEAE-dextran, 66 µM chloroquine, and 4 µg DNA in serum free DME. For each set, a positive control is prepared, e.g., of DTLR-FLAG cDNA at 1 and 1/200 dilution, and a negative mock. Rinse cells with 35 serum free DME. Add the DNA solution and incubate 5 hr at 37° C. Remove the medium and add 0.5 ml 10% DMSO in

DME for 2.5 min. Remove and wash once with DME. Add 1.5 ml growth medium and incubate overnight.

On day 2, change the medium. On days 3 or 4, the cells are fixed and stained. Rinse the cells twice with

- 5 Hank's Buffered Saline Solution (HBSS) and fix in 4% paraformaldehyde (PFA)/glucose for 5 min. Wash 3X with HBSS. The slides may be stored at -80° C after all liquid is removed. For each chamber, 0.5 ml incubations are performed as follows. Add HBSS/saponin (0.1%) with 32 µl/ml of 1 M NaN<sub>3</sub> for 20 min. Cells are then washed with HBSS/saponin 1X. Add appropriate DTLR or DTLR/antibody complex to cells and incubate for 30 min. Wash cells twice with HBSS/saponin. If appropriate, add first antibody for 30 min. Add second antibody, e.g.,
- 10 Vector anti-mouse antibody, at 1/200 dilution, and incubate for 30 min. Prepare ELISA solution, e.g., Vector Elite ABC horseradish peroxidase solution, and preincubate for 30 min. Use, e.g., 1 drop of solution A (avidin) and 1 drop solution B (biotin) per 2.5 ml
- 15 HBSS/saponin. Wash cells twice with HBSS/saponin. Add ABC HRP solution and incubate for 30 min. Wash cells twice with HBSS, second wash for 2 min, which closes cells. Then add Vector diaminobenzoic acid (DAB) for 5 to 10 min. Use 2 drops of buffer plus 4 drops DAB plus 2
- 20 drops of H<sub>2</sub>O<sub>2</sub> per 5 ml of glass distilled water.
- 25 Carefully remove chamber and rinse slide in water. Air dry for a few minutes, then add 1 drop of Crystal Mount and a cover slip. Bake for 5 min at 85-90° C.

- Evaluate positive staining of pools and
- 30 progressively subclone to isolation of single genes responsible for the binding.

Alternatively, DTLR reagents are used to affinity purify or sort out cells expressing a putative ligand. See, e.g., Sambrook, et al. or Ausubel, et al.

- 35 Another strategy is to screen for a membrane bound receptor by panning. The receptor cDNA is constructed as described above. The ligand can be immobilized and used

to immobilize expressing cells. Immobilization may be achieved by use of appropriate antibodies which recognize, e.g., a FLAG sequence of a DTLR fusion construct, or by use of antibodies raised against the 5 first antibodies. Recursive cycles of selection and amplification lead to enrichment of appropriate clones and eventual isolation of receptor expressing clones.

Phage expression libraries can be screened by mammalian DTLRs. Appropriate label techniques, e.g., 10 anti-FLAG antibodies, will allow specific labeling of appropriate clones.

All citations herein are incorporated herein by reference to the same extent as if each individual publication or patent 15 application was specifically and individually indicated to be incorporated by reference.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be 20 apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited by the terms of the appended claims, along with the full scope of equivalents to which such claims 25 are entitled; and the invention is not to be limited by the specific embodiments that have been presented herein by way of example.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

(i) APPLICANT: (A) NAME: Schering Corporation  
(B) STREET: 2000 Galloping Hill Road  
(C) CITY: Kenilworth  
(D) STATE: New Jersey  
10 (E) COUNTRY: USA  
(F) POSTAL CODE: 07033  
(G) TELEPHONE: (908) 298-4000  
(H) TELEFAX: (908) 298-5388

15 (ii) TITLE OF INVENTION: HUMAN RECEPTOR PROTEINS; RELATED REAGENTS AND METHODS

(iii) NUMBER OF SEQUENCES: 35

20

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: Macintosh Power PC  
(C) OPERATING SYSTEM: 8.0  
25 (D) SOFTWARE: Microsoft Word 6.0

(v) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

30

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NO.: USSN 60/044,293  
35 (B) FILING DATE: 07-MAY-1997

35

(A) APPLICATION NO.: USSN 60/072,212  
(B) FILING DATE: 22-JAN-1998

(A) APPLICATION NO.: USSN 60/076,947  
40 (B) FILING DATE: 05-MAR-1998

## (2) INFORMATION FOR SEQ ID NO:1:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2367 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
50 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50

(ix) FEATURE:

(A) NAME/KEY: CDS  
55 (B) LOCATION: 1..2358

(ix) FEATURE:

(A) NAME/KEY: mat\_peptide  
(B) LOCATION: 67..2358

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	ATG ACT AGC ATC TTC CAT TTT GCC ATT ATC TTC ATG TTA ATA CTT CAG	48
5	Met Thr Ser Ile Phe His Phe Ala Ile Ile Phe Met Leu Ile Leu Gln	
	-22 -20 -15 -10	
	ATC AGA ATA CAA TTA TCT GAA GAA AGT GAA TTT TTA GTT GAT AGG TCA	96
10	Ile Arg Ile Gln Leu Ser Glu Glu Ser Glu Phe Leu Val Asp Arg Ser	
	-5 1 5 10.	
	AAA AAC GGT CTC ATC CAC GTT CCT AAA GAC CTA TCC CAG AAA ACA ACA	144
	Lys Asn Gly Leu Ile His Val Pro Lys Asp Leu Ser Gln Lys Thr Thr	
	15 20 25	
15	ATC TTA AAT ATA TCG CAA AAT TAT ATA TCT GAG CTT TGG ACT TCT GAC	192
	Ile Leu Asn Ile Ser Gln Asn Tyr Ile Ser Glu Leu Trp Thr Ser Asp	
	30 35 40	
20	ATC TTA TCA CTG TCA AAA CTG AGG ATT TTG ATA ATT TCT CAT AAT AGA	240
	Ile Leu Ser Leu Ser Lys Leu Arg Ile Leu Ile Ser His Asn Arg	
	45 50 55	
25	ATC CAG TAT CTT GAT ATC AGT GTT TTC AAA TTC AAC CAG GAA TTG GAA	288
	Ile Gln Tyr Leu Asp Ile Ser Val Phe Lys Phe Asn Gln Glu Leu Glu	
	60 65 70	
	TAC TTG GAT TTG TCC CAC AAC AAG TTG GTG AAG ATT TCT TGC CAC CCT	336
30	Tyr Leu Asp Leu Ser His Asn Lys Leu Val Lys Ile Ser Cys His Pro	
	75 80 85 90	
	ACT GTG AAC CTC AAG CAC TTG GAC CTG TCA TTT AAT GCA TTT GAT GCC	384
	Thr Val Asn Leu Lys His Leu Asp Leu Ser Phe Asn Ala Phe Asp Ala	
	95 100 105	
35	CTG CCT ATA TGC AAA GAG TTT GGC AAT ATG TCT CAA CTA AAA TTT CTG	432
	Leu Pro Ile Cys Lys Glu Phe Gly Asn Met Ser Gln Leu Lys Phe Leu	
	110 115 120	
40	GGG TTG AGC ACC ACA CAC TTA GAA AAA TCT AGT GTG CTG CCA ATT GCT	480
	Gly Leu Ser Thr His Leu Glu Lys Ser Ser Val Leu Pro Ile Ala	
	125 130 135	
45	CAT TTG AAT ATC AGC AAG GTC TTG CTG GTC TTA GGA GAG ACT TAT GGG	528
	His Leu Asn Ile Ser Lys Val Leu Leu Val Leu Gly Glu Thr Tyr Gly	
	140 145 150	
	GAA AAA GAA GAC CCT GAG GGC CTT CAA GAC TTT AAC ACT GAG AGT CTG	576
50	Glu Lys Glu Asp Pro Glu Gly Leu Gln Asp Phe Asn Thr Glu Ser Leu	
	155 160 165 170	
	CAC ATT GTG TTC CCC ACA AAC AAA GAA TTC CAT TTT ATT TTG GAT GTG	624
	His Ile Val Phe Pro Thr Asn Lys Glu Phe His Phe Ile Leu Asp Val	
	175 180 185	
55	TCA GTC AAG ACT GTA GCA AAT CTG GAA CTA TCT AAT ATC AAA TGT GTG	672
	Ser Val Lys Thr Val Ala Asn Leu Glu Leu Ser Asn Ile Lys Cys Val	
	190 195 200	
60	CTA GAA GAT AAC AAA TGT TCT TAC TTC CTA AGT ATT CTG GCG AAA CTT	720
	Leu Glu Asp Asn Lys Cys Ser Tyr Phe Leu Ser Ile Leu Ala Lys Leu	

	205	210	215	
5	CAA ACA AAT CCA AAG TTA TCA AGT CTT ACC TTA AAC AAC ATT GAA ACA Gln Thr Asn Pro Lys Leu Ser Ser Leu Thr Leu Asn Asn Ile Glu Thr 220 225 230			768
10	ACT TGG AAT TCT TTC ATT AGG ATC CTC CAA CTA GTT TGG CAT ACA ACT Thr Trp Asn Ser Phe Ile Arg Ile Leu Gln Leu Val Trp His Thr Thr 235 240 245 250			816
15	GTA TGG TAT TTC TCA ATT TCA AAC GTG AAG CTA CAG GGT CAG CTG GAC Val Trp Tyr Phe Ser Ile Ser Asn Val Lys Leu Gln Gly Gln Leu Asp 255 260 265			864
20	TTC AGA GAT TTT GAT TAT TCT GGC ACT TCC TTG AAG GCC TTG TCT ATA Phe Arg Asp Phe Asp Tyr Ser Gly Thr Ser Leu Lys Ala Leu Ser Ile 270 275 280			912
25	CAC CAA GTT GTC AGC GAT GTG TTC GGT TTT CCG CAA AGT TAT ATC TAT His Gln Val Val Ser Asp Val Phe Gly Phe Pro Gln Ser Tyr Ile Tyr 285 290 295			960
30	GAA ATC TTT TCG AAT ATG AAC ATC AAA AAT TTC ACA GTG TCT GGT ACA Glu Ile Phe Ser Asn Met Asn Ile Lys Asn Phe Thr Val Ser Gly Thr 300 305 310			1008
35	CGC ATG GTC CAC ATG CTT TGC CCA TCC AAA ATT AGC CCG TTC CTG CAT Arg Met Val His Met Leu Cys Pro Ser Lys Ile Ser Pro Phe Leu His 315 320 325 330			1056
40	TTG GAT TTT TCC AAT AAT CTC TTA ACA GAC ACG GTT TTT GAA AAT TGT Leu Asp Phe Ser Asn Asn Leu Leu Thr Asp Thr Val Phe Glu Asn Cys 335 340 345			1104
45	GGG CAC CTT ACT GAG TTG GAG ACA CTT ATT TTA CAA ATG AAT CAA TTA Gly His Leu Thr Glu Leu Glu Thr Leu Ile Leu Gln Met Asn Gln Leu 350 355 360			1152
50	AAA GAA CTT TCA AAA ATA GCT GAA ATG ACT ACA CAG ATG AAG TCT CTG Lys Glu Leu Ser Lys Ile Ala Glu Met Thr Thr Gln Met Lys Ser Leu 365 370 375			1200
55	CAA CAA TTG GAT ATT AGC CAG AAT TCT GTA AGC TAT GAT GAA AAG AAA Gln Gln Leu Asp Ile Ser Gln Asn Ser Val Ser Tyr Asp Glu Lys Lys 380 385 390			1248
60	GGA GAC TGT TCT TGG ACT AAA AGT TTA TTA AGT TTA AAT ATG TCT TCA Gly Asp Cys Ser Trp Thr Lys Ser Leu Leu Ser Leu Asn Met Ser Ser 395 400 405 410			1296
	AAT ATA CTT ACT GAC ACT ATT TTC AGA TGT TTA CCT CCC AGG ATC AAG Asn Ile Leu Thr Asp Thr Ile Phe Arg Cys Leu Pro Pro Arg Ile Lys 415 420 425			1344
	GTA CTT GAT CTT CAC AGC AAT AAA ATA AAG AGC ATT CCT AAA CAA GTC Val Leu Asp Leu His Ser Asn Lys Ile Lys Ser Ile Pro Lys Gln Val 430 435 440			1392
	GTA AAA CTG GAA GCT TTG CAA GAA CTC AAT GTT GCT TTC AAT TCT TTA Val Lys Leu Glu Ala Leu Gln Glu Leu Asn Val Ala Phe Asn Ser Leu 445 450 455			1440

	ACT GAC CTT CCT GGA TGT GGC AGC TTT AGC AGC CTT TCT GTA TTG ATC Thr Asp Leu Pro Gly Cys Gly Ser Phe Ser Ser Leu Ser Val Leu Ile 460 465 470	1488
5	ATT GAT CAC AAT TCA GTT TCC CAC CCA TCA GCT GAT TTC TTC CAG AGC Ile Asp His Asn Ser Val Ser His Pro Ser Ala Asp Phe Phe Gln Ser 475 480 485 490	1536
10	TGC CAG AAG ATG AGG TCA ATA AAA GCA GGG GAC AAT CCA TTC CAA TGT. Cys Gln Lys Met Arg Ser Ile Lys Ala Gly Asp Asn Pro Phe Gln Cys 495 500 505	1584
15	ACC TGT GAG CTC GGA GAA TTT GTC AAA AAT ATA GAC CAA GTA TCA AGT Thr Cys Glu Leu Gly Glu Phe Val Lys Asn Ile Asp Gln Val Ser Ser 510 515 520	1632
20	GAA GTG TTA GAG GGC TGG CCT GAT TCT TAT AAG TGT GAC TAC CCG GAA Glu Val Leu Glu Gly Trp Pro Asp Ser Tyr Lys Cys Asp Tyr Pro Glu 525 530 535	1680
25	AGT TAT AGA GGA ACC CTA CTA AAG GAC TTT CAC ATG TCT GAA TTA TCC Ser Tyr Arg Gly Thr Leu Leu Lys Asp Phe His Met Ser Glu Leu Ser 540 545 550	1728
30	TGC AAC ATA ACT CTG CTG ATC GTC ACC ATC GTT GCC ACC ATG CTG GTG Cys Asn Ile Thr Leu Leu Ile Val Thr Ile Val Ala Thr Met Leu Val 555 560 565 570	1776
35	TTG GCT GTG ACT GTG ACC TCC CTC TGC ATC TAC TTG GAT CTG CCC TGG Leu Ala Val Thr Val Thr Ser Leu Cys Ile Tyr Leu Asp Leu Pro Trp 575 580 585	1824
40	TAT CTC AGG ATG GTG TGC CAG TGG ACC ACC CAG ACC CGG CGC AGG GCC AGG Tyr Leu Arg Met Val Cys Gln Trp Thr Gln Thr Arg Arg Arg Ala Arg 590 595 600	1872
45	AAC ATA CCC TTA GAA GAA CTC CAA AGA AAT CTC CAG TTT CAT GCA TTT Asn Ile Pro Leu Glu Glu Leu Gln Arg Asn Leu Gln Phe His Ala Phe 605 610 615	1920
50	ATT TCA TAT AGT GGG CAC GAT TCT TTC TGG GTG AAG AAT GAA TTA TTG Ile Ser Tyr Ser Gly His Asp Ser Phe Trp Val Lys Asn Glu Leu Leu 620 625 630	1968
55	CCA AAC CTA GAG AAA GAA GGT ATG CAG ATT TGC CTT CAT GAG AGA AAC Pro Asn Leu Glu Lys Glu Gly Met Gln Ile Cys Leu His Glu Arg Asn 635 640 645 650	2016
60	TTT GTT CCT GGC AAG AGC ATT GTG GAA AAT ATC ATC ACC TGC ATT GAG Phe Val Pro Gly Lys Ser Ile Val Glu Asn Ile Ile Thr Cys Ile Glu 655 660 665	2064
65	AAG AGT TAC AAG TCC ATC TTT GTT TTG TCT CCC AAC TTT GTC CAG AGT Lys Ser Tyr Lys Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Ser 670 675 680	2112
70	GAA TGG TGC CAT TAT GAA CTC TAC TTT GCC CAT CAC AAT CTC TTT CAT Glu Trp Cys His Tyr Glu Leu Tyr Phe Ala His His Asn Leu Phe His 685 690 695	2160

	GAA GGA TCT AAT AGC TTA ATC CTG ATC TTG CTG GAA CCC ATT CCG CAG Glu Gly Ser Asn Ser Leu Ile Leu Ile Leu Leu Glu Pro Ile Pro Gln 700                   705                   710	2208
5	TAC TCC ATT CCT AGC AGT TAT CAC AAG CTC AAA AGT CTC ATG GCC AGG Tyr Ser Ile Pro Ser Ser Tyr His Lys Leu Lys Ser Leu Met Ala Arg 715                   720                   725                   730	2256
10	AGG ACT TAT TTG GAA TGG CCC AAG GAA AAG AGC AAA CGT GGC CTT TTT Arg Thr Tyr Leu Glu Trp Pro Lys Glu Lys Ser Lys Arg Gly Leu Phe 735                   740                   745	2304
15	TGG GCT AAC TTA AGG GCA GCC ATT AAT ATT AAG CTG ACA GAG CAA GCA Trp Ala Asn Leu Arg Ala Ala Ile Asn Ile Lys Leu Thr Glu Gln Ala 750                   755                   760	2352
20	AAG AAA TAGTCTAGA Lys Lys	2367

## (2) INFORMATION FOR SEQ ID NO:2:

25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 786 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear
30	(ii) MOLECULE TYPE: protein
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
40	Met Thr Ser Ile Phe His Phe Ala Ile Ile Phe Met Leu Ile Leu Gln -22       -20                   -15                   -10
45	Ile Arg Ile Gln Leu Ser Glu Glu Ser Glu Phe Leu Val Asp Arg Ser -5                   1                   5                   10
50	Lys Asn Gly Leu Ile His Val Pro Lys Asp Leu Ser Gln Lys Thr Thr 15                   20                   25
55	Ile Leu Asn Ile Ser Gln Asn Tyr Ile Ser Glu Leu Trp Thr Ser Asp 30                   35                   40
60	Ile Leu Ser Leu Ser Lys Leu Arg Ile Leu Ile Ile Ser His Asn Arg 45                   50                   55
65	Ile Gln Tyr Leu Asp Ile Ser Val Phe Lys Phe Asn Gln Glu Leu Glu 60                   65                   70
70	Tyr Leu Asp Leu Ser His Asn Lys Leu Val Lys Ile Ser Cys His Pro 75                   80                   85                   90
75	Thr Val Asn Leu Lys His Leu Asp Leu Ser Phe Asn Ala Phe Asp Ala 95                   100                   105
80	Leu Pro Ile Cys Lys Glu Phe Gly Asn Met Ser Gln Leu Lys Phe Leu 110                   115                   120
85	Gly Leu Ser Thr Thr His Leu Glu Lys Ser Ser Val Leu Pro Ile Ala 125                   130                   135

	His Leu Asn Ile Ser Lys Val Leu Leu Val Leu Gly Glu Thr Tyr Gly			
	140	145	150	
5	Glu Lys Glu Asp Pro Glu Gly Leu Gln Asp Phe Asn Thr Glu Ser Leu			
	155	160	165	170
	His Ile Val Phe Pro Thr Asn Lys Glu Phe His Phe Ile Leu Asp Val			
	175	180	185	
10	Ser Val Lys Thr Val Ala Asn Leu Glu Leu Ser Asn Ile Lys Cys Val			
	190	195	200	
15	Leu Glu Asp Asn Lys Cys Ser Tyr Phe Leu Ser Ile Leu Ala Lys Leu			
	205	210	215	
	Gln Thr Asn Pro Lys Leu Ser Ser Leu Thr Leu Asn Asn Ile Glu Thr			
	220	225	230	
20	Thr Trp Asn Ser Phe Ile Arg Ile Leu Gln Leu Val Trp His Thr Thr			
	235	240	245	250
	Val Trp Tyr Phe Ser Ile Ser Asn Val Lys Leu Gln Gly Gln Leu Asp			
	255	260	265	
25	Phe Arg Asp Phe Asp Tyr Ser Gly Thr Ser Leu Lys Ala Leu Ser Ile			
	270	275	280	
30	His Gln Val Val Ser Asp Val Phe Gly Phe Pro Gln Ser Tyr Ile Tyr			
	285	290	295	
	Glu Ile Phe Ser Asn Met Asn Ile Lys Asn Phe Thr Val Ser Gly Thr			
	300	305	310	
35	Arg Met Val His Met Leu Cys Pro Ser Lys Ile Ser Pro Phe Leu His			
	315	320	325	330
	Leu Asp Phe Ser Asn Asn Leu Leu Thr Asp Thr Val Phe Glu Asn Cys			
	335	340	345	
40	Gly His Leu Thr Glu Leu Glu Thr Leu Ile Leu Gln Met Asn Gln Leu			
	350	355	360	
45	Lys Glu Leu Ser Lys Ile Ala Glu Met Thr Thr Gln Met Lys Ser Leu			
	365	370	375	
	Gln Gln Leu Asp Ile Ser Gln Asn Ser Val Ser Tyr Asp Glu Lys Lys			
	380	385	390	
50	Gly Asp Cys Ser Trp Thr Lys Ser Leu Leu Ser Leu Asn Met Ser Ser			
	395	400	405	410
	Asn Ile Leu Thr Asp Thr Ile Phe Arg Cys Leu Pro Pro Arg Ile Lys			
	415	420	425	
55	Val Leu Asp Leu His Ser Asn Lys Ile Lys Ser Ile Pro Lys Gln Val			
	430	435	440	
60	Val Lys Leu Glu Ala Leu Gln Glu Leu Asn Val Ala Phe Asn Ser Leu			
	445	450	455	

Thr Asp Leu Pro Gly Cys Gly Ser Phe Ser Ser Leu Ser Val Leu Ile  
 460 465 470

5 Ile Asp His Asn Ser Val Ser His Pro Ser Ala Asp Phe Phe Gln Ser  
 475 480 485 490

Cys Gln Lys Met Arg Ser Ile Lys Ala Gly Asp Asn Pro Phe Gln Cys  
 495 500 505

10 Thr Cys Glu Leu Gly Glu Phe Val Lys Asn Ile Asp Gln Val Ser Ser.  
 510 515 520

Glu Val Leu Glu Gly Trp Pro Asp Ser Tyr Lys Cys Asp Tyr Pro Glu  
 525 530 535

15 Ser Tyr Arg Gly Thr Leu Leu Lys Asp Phe His Met Ser Glu Leu Ser  
 540 545 550

20 Cys Asn Ile Thr Leu Leu Ile Val Thr Ile Val Ala Thr Met Leu Val  
 555 560 565 570

Leu Ala Val Thr Val Thr Ser Leu Cys Ile Tyr Leu Asp Leu Pro Trp  
 575 580 585

25 Tyr Leu Arg Met Val Cys Gln Trp Thr Gln Thr Arg Arg Arg Ala Arg  
 590 595 600

Asn Ile Pro Leu Glu Leu Gln Arg Asn Leu Gln Phe His Ala Phe  
 605 610 615

30 Ile Ser Tyr Ser Gly His Asp Ser Phe Trp Val Lys Asn Glu Leu Leu  
 620 625 630

35 Pro Asn Leu Glu Lys Glu Gly Met Gln Ile Cys Leu His Glu Arg Asn  
 635 640 645 650

Phe Val Pro Gly Lys Ser Ile Val Glu Asn Ile Ile Thr Cys Ile Glu  
 655 660 665

40 Lys Ser Tyr Lys Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Ser  
 670 675 680

Glu Trp Cys His Tyr Glu Leu Tyr Phe Ala His His Asn Leu Phe His  
 685 690 695

45 Glu Gly Ser Asn Ser Leu Ile Leu Ile Leu Glu Pro Ile Pro Gln  
 700 705 710

50 Tyr Ser Ile Pro Ser Ser Tyr His Lys Leu Lys Ser Leu Met Ala Arg  
 715 720 725 730

Arg Thr Tyr Leu Glu Trp Pro Lys Glu Lys Ser Lys Arg Gly Leu Phe  
 735 740 745

55 Trp Ala Asn Leu Arg Ala Ala Ile Asn Ile Lys Leu Thr Glu Gln Ala  
 750 755 760

Lys Lys

(ii) MOLECULE TYPE: cDNA

10  
(ix) FEATURE:  
      (A) NAME/KEY: CDS

15 (ix) FEATURE:  
· (A) NAME/KEY: mat\_peptide  
· (B) LOCATION: 67 2352

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

	ATG CCA CAT ACT TTG TGG ATG GTG TGG GTC TTG GGG GTC ATC ATC AGC	48
	Met Pro His Thr Leu Trp Met Val Trp Val Leu Gly Val Ile Ile Ser	
-22	-20	-15
-10		
25		
	CTC TCC AAG GAA GAA TCC TCC AAT CAG GCT TCT CTG TCT TGT GAC CGC	96
	Leu Ser Lys Glu Glu Ser Ser Asn Gln Ala Ser Leu Ser Cys Asp Arg	
-5	1	5
		10
30		
	AAT GGT ATC TGC AAG GGC AGC TCA GGA TCT TTA AAC TCC ATT CCC TCA	144
	Asn Gly Ile Cys Lys Gly Ser Ser Gly Ser Leu Asn Ser Ile Pro Ser	
15	20	25
35		
	GGG CTC ACA GAA GCT GTA AAA AGC CTT GAC CTG TCC AAC AAC AGG ATC	192
	Gly Leu Thr Glu Ala Val Lys Ser Leu Asp Leu Ser Asn Asn Arg Ile	
30	35	40
40		
	ACC TAC ATT AGC AAC AGT GAC CTA CAG AGG TGT GTG AAC CTC CAG GCT	240
	Thr Tyr Ile Ser Asn Ser Asp Leu Gln Arg Cys Val Asn Leu Gln Ala	
45	45	50
		55
45		
	CTG GTG CTG ACA TCC AAT GGA ATT AAC ACA ATA GAG GAA GAT TCT TTT	288
	Leu Val Leu Thr Ser Asn Gly Ile Asn Thr Ile Glu Glu Asp Ser Phe	
60	65	70
50		
	TCT TCC CTG GGC AGT CTT GAA CAT TTA GAC TTA TCC TAT AAT TAC TTA	336
	Ser Ser Leu Gly Ser Leu Glu His Leu Asp Leu Ser Tyr Asn Tyr Leu	
75	80	85
		90
55		
	TCT AAT TTA TCG TCT TCC TGG TTC AAG CCC CTT TCT TCT TTA ACA TTC	384
	Ser Asn Leu Ser Ser Ser Trp Phe Lys Pro Leu Ser Ser Leu Thr Phe	
95	100	105
55		
	TTA AAC TTA CTG GGA AAT CCT TAC AAA ACC CTA GGG GAA ACA TCT CTT	432
	Leu Asn Leu Leu Gly Asn Pro Tyr Lys Thr Leu Gly Glu Thr Ser Leu	
110	115	120
60		
	TTT TCT CAT CTC ACA AAA TTG CAA ATC CTG AGA GTG GGA AAT ATG GAC	480
	Phe Ser His Leu Thr Lys Leu Gln Ile Leu Arg Val Gly Asn Met Asp	
125	130	135

	ACC TTC ACT AAG ATT CAA AGA AAA GAT TTT GCT GGA CTT ACC TTC CTT Thr Phe Thr Lys Ile Gln Arg Lys Asp Phe Ala Gly Leu Thr Phe Leu 140 145 150	528
5	GAG GAA CTT GAG ATT GAT GCT TCA GAT CTA CAG AGC TAT GAG CCA AAA Glu Glu Leu Glu Ile Asp Ala Ser Asp Leu Gln Ser Tyr Glu Pro Lys 155 160 165 170	576
10	AGT TTG AAG TCA ATT CAG AAC GTA AGT CAT CTG ATC CTT CAT ATG AAG Ser Leu Lys Ser Ile Gln Asn Val Ser His Leu Ile Leu His Met Lys 175 180 185	624
15	CAG CAT ATT TTA CTG CTG GAG ATT TTT GTA GAT GTT ACA AGT TCC GTG Gln His Ile Leu Leu Glu Ile Phe Val Asp Val Thr Ser Ser Val 190 195 200	672
20	GAA TGT TTG GAA CTG CGA GAT ACT GAT TTG GAC ACT TTC CAT TTT TCA Glu Cys Leu Glu Leu Arg Asp Thr Asp Leu Asp Thr Phe His Phe Ser 205 210 215	720
25	GAA CTA TCC ACT GGT GAA ACA AAT TCA TTG ATT AAA AAG TTT ACA TTT Glu Leu Ser Thr Gly Glu Thr Asn Ser Leu Ile Lys Lys Phe Thr Phe 220 225 230	768
30	AGA AAT GTG AAA ATC ACC GAT GAA AGT TTG TTT CAG GTT ATG AAA CTT Arg Asn Val Lys Ile Thr Asp Glu Ser Leu Phe Gln Val Met Lys Leu 235 240 245 250	816
35	TTG AAT CAG ATT TCT GGA TTG TTA GAA TTA GAG TTT GAT GAC TGT ACC Leu Asn Gln Ile Ser Gly Leu Leu Glu Leu Glu Phe Asp Asp Cys Thr 255 260 265	864
40	CTT AAT GGA GTT GGT AAT TTT AGA GCA TCT GAT AAT GAC AGA GTT ATA Leu Asn Gly Val Gly Asn Phe Arg Ala Ser Asp Asn Asp Arg Val Ile 270 275 280	912
45	GAT CCA GGT AAA GTG GAA ACG TTA ACA ATC CGG AGG CTG CAT ATT CCA Asp Pro Gly Lys Val Glu Thr Leu Thr Ile Arg Arg Leu His Ile Pro 285 290 295	960
50	AGG TTT TAC TTA TTT TAT GAT CTG AGC ACT TTA TAT TCA CTT ACA GAA Arg Phe Tyr Leu Phe Tyr Asp Leu Ser Thr Leu Tyr Ser Leu Thr Glu 300 305 310	1008
55	AGA GTT AAA AGA ATC ACA GTA GAA AAC AGT AAA GTT TTT CTG GTT CCT Arg Val Lys Arg Ile Thr Val Glu Asn Ser Lys Val Phe Leu Val Pro 315 320 325 330	1056
60	TGT TTA CTT TCA CAA CAT TTA AAA TCA TTA GAA TAC TTG GAT CTC AGT Cys Leu Leu Ser Gln His Leu Lys Ser Leu Glu Tyr Leu Asp Leu Ser 335 340 345	1104
	GAA AAT TTG ATG GTT GAA GAA TAC TTG AAA AAT TCA GCC TGT GAG GAT Glu Asn Leu Met Val Glu Glu Tyr Leu Lys Asn Ser Ala Cys Glu Asp 350 355 360	1152
	GCC TGG CCC TCT CTA CAA ACT TTA ATT TTA AGG CAA AAT CAT TTG GCA Ala Trp Pro Ser Leu Gln Thr Leu Ile Leu Arg Gln Asn His Leu Ala 365 370 375	1200
	TCA TTG GAA AAA ACC GGA GAG ACT TTG CTC ACT CTG AAA AAC TTG ACT	1248

	Ser Leu Glu Lys Thr Gly Glu Thr Leu Leu Thr Leu Lys Asn Leu Thr			
	380	385	390	
5	AAC ATT GAT ATC AGT AAG AAT AGT TTT CAT TCT ATG CCT GAA ACT TGT Asn Ile Asp Ile Ser Lys Asn Ser Phe His Ser Met Pro Glu Thr Cys		1296	
	395	400	405	410
10	CAG TGG CCA GAA AAG ATG AAA TAT TTG AAC TTA TCC AGC ACA CGA ATA Gln Trp Pro Glu Lys Met Lys Tyr Leu Asn Leu Ser Ser Thr Arg Ile		1344	
	415	420	425	
15	CAC AGT GTA ACA GGC TGC ATT CCC AAG ACA CTG GAA ATT TTA GAT GTT His Ser Val Thr Gly Cys Ile Pro Lys Thr Leu Glu Ile Leu Asp Val		1392	
	430	435	440	
20	AGC AAC AAC AAT CTC AAT TTA TTT TCT TTG AAT TTG CCG CAA CTC AAA Ser Asn Asn Asn Leu Asn Leu Phe Ser Leu Asn Leu Pro Gln Leu Lys		1440	
	445	450	455	
25	GAA CTT TAT ATT TCC AGA AAT AAG TTG ATG ACT CTA CCA GAT GCC TCC Glu Leu Tyr Ile Ser Arg Asn Lys Leu Met Thr Leu Pro Asp Ala Ser		1488	
	460	465	470	
30	CTC TTA CCC ATG TTA CT <sub>A</sub> GTA TTG AAA ATC AGT AGG AAT GCA ATA ACT Leu Leu Pro Met Leu Leu Val Leu Lys Ile Ser Arg Asn Ala Ile Thr		1536	
	475	480	485	490
35	ACG TTT TCT AAG GAG CAA CTT GAC TCA TTT CAC ACA CTG AAG ACT TTG Thr Phe Ser Lys Glu Gln Leu Asp Ser Phe His Thr Leu Lys Thr Leu		1584	
	495	500	505	
40	GAA GCT GGT GGC AAT AAC TTC ATT TGC TCC TGT GAA TTC CTC TCC TTC Glu Ala Gly Gly Asn Asn Phe Ile Cys Ser Cys Glu Phe Leu Ser Phe		1632	
	510	515	520	
45	525	530	535	
50	ACT CAG GAG CAG CAA GCA CTG GCC AAA GTC TTG ATT GAT TGG CCA GCA Thr Gln Glu Gln Gln Ala Leu Ala Lys Val Leu Ile Asp Trp Pro Ala		1680	
	540	545	550	
55	AAT TAC CTG TGT GAC TCT CCA TCC CAT GTG CGT GGC CAG CAG GTT CAG Asn Tyr Leu Cys Asp Ser Pro Ser His Val Arg Gly Gln Gln Val Gln		1728	
	555	560	565	570
60	GAT GTC CGC CTC TCG GTG TCG GAA TGT CAC AGG ACA GCA CTG GTG TCT Asp Val Arg Leu Ser Val Ser Glu Cys His Arg Thr Ala Leu Val Ser		1776	
	575	580	585	
	590	595	600	
65	GGC ATG TGC TGT GCT CTG TTC CTG CTG ATC CTG CTC ACG GGG GTC CTG Gly Met Cys Cys Ala Leu Phe Leu Leu Ile Leu Leu Thr Gly Val Leu		1824	
	605	610	615	
70	TGC CAC CGT TTC CAT GGC CTG TGG TAT ATG AAA ATG ATG TGG GCC TGG Cys His Arg Phe His Gly Leu Trp Tyr Met Lys Met Met Trp Ala Trp		1872	
	620	625	630	
75	CTC CAG GCC AAA AGG AAG CCC AGG AAA GCT CCC AGC AGG AAC ATC TGC Leu Gln Ala Lys Arg Lys Pro Arg Lys Ala Pro Ser Arg Asn Ile Cys		1920	
	640	645	650	
80	TAT GAT GCA TTT GTT TCT TAC AGT GAG CGG GAT GCC TAC TGG GTG GAG Tyr Asp Ala Phe Val Ser Tyr Ser Glu Arg Asp Ala Tyr Trp Val Glu		1968	

	620	625	630	
	AAC CTT ATG GTC CAG GAG CTG GAG AAC TTC AAT CCC CCC TTC AAG TTG			2016
5	Asn Leu Met Val Gln Glu Leu Glu Asn Phe Asn Pro Pro Phe Lys Leu			
	635	640	645	650
	TGT CTT CAT AAG CGG GAC TTC ATT CCT GGC AAG TGG ATC ATT GAC AAT			2064
	Cys Leu His Lys Arg Asp Phe Ile Pro Gly Lys Trp Ile Ile Asp Asn			
10	655		660	665
	ATC ATT GAC TCC ATT GAA AAG AGC CAC AAA ACT GTC TTT GTG CTT TCT			2112
	Ile Ile Asp Ser Ile Glu Lys Ser His Lys Thr Val Phe Val Leu Ser			
	670		675	680
15	GAA AAC TTT GTG AAG AGT GAG TGG TGC AAG TAT GAA CTG GAC TTC TCC			2160
	Glu Asn Phe Val Lys Ser Glu Trp Cys Lys Tyr Glu Leu Asp Phe Ser			
	685	690		695
20	CAT TTC CGT CTT TTT GAA GAG AAC AAT GAT GCT GCC ATT CTC ATT CTT			2208
	His Phe Arg Leu Phe Glu Glu Asn Asn Asp Ala Ala Ile Leu Ile Leu			
	700	705		710
	CTG GAG CCC ATT GAG AAA AAA GCC ATT CCC CAG CGC TTC TGC AAG CTG			2256
	Leu Glu Pro Ile Glu Lys Lys Ala Ile Pro Gln Arg Phe Cys Lys Leu			
25	715	720	725	730
	CGG AAG ATA ATG AAC ACC AAG ACC TAC CTG GAG TGG CCC ATG GAC GAG			2304
	Arg Lys Ile Met Asn Thr Lys Thr Tyr Leu Glu Trp Pro Met Asp Glu			
	735		740	745
30	GCT CAG CGG GAA GGA TTT TGG GTA AAT CTG AGA GCT GCG ATA AAG TCC			2352
	Ala Gln Arg Glu Gly Phe Trp Val Asn Leu Arg Ala Ala Ile Lys Ser			
	750		755	760
35	TAG			2355

(2) INFORMATION FOR SEQ ID NO:4:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 784 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

50	Met	Pro	His	Thr	Leu	Trp	Met	Val	Trp	Val	Leu	Gly	Val	Ile	Ile	Ser
	-22								-15							-10
	Leu	Ser	Lys	Glu	Glu	Ser	Ser	Asn	Gln	Ala	Ser	Leu	Ser	Cys	Asp	Arg
55	Asn	Gly	Ile	Cys	Lys	Gly	Ser	Ser	Gly	Ser	Leu	Asn	Ser	Ile	Pro	Ser
	Gly	Leu	Thr	Glu	Ala	Val	Lys	Ser	Leu	Asp	Leu	Ser	Asn	Asn	Arg	Ile
60																
	Thr	Tyr	Ile	Ser	Asn	Ser	Asp	Leu	Gln	Arg	Cys	Val	Asn	Leu	Gln	Ala

	45	50	55
	Leu Val Leu Thr Ser Asn Gly Ile Asn Thr Ile Glu Glu Asp Ser Phe		
	60	65	70
5	Ser Ser Leu Gly Ser Leu Glu His Leu Asp Leu Ser Tyr Asn Tyr Leu		
	75	80	85
	90		
10	Ser Asn Leu Ser Ser Ser Trp Phe Lys Pro Leu Ser Ser Leu Thr Phe		
	95	100	105
	Leu Asn Leu Leu Gly Asn Pro Tyr Lys Thr Leu Gly Glu Thr Ser Leu		
	110	115	120
15	Phe Ser His Leu Thr Lys Leu Gln Ile Leu Arg Val Gly Asn Met Asp		
	125	130	135
	Thr Phe Thr Lys Ile Gln Arg Lys Asp Phe Ala Gly Leu Thr Phe Leu		
	140	145	150
20	Glu Glu Leu Glu Ile Asp Ala Ser Asp Leu Gln Ser Tyr Glu Pro Lys		
	155	160	165
	170		
25	Ser Leu Lys Ser Ile Gln Asn Val Ser His Leu Ile Leu His Met Lys		
	175	180	185
	Gln His Ile Leu Leu Leu Glu Ile Phe Val Asp Val Thr Ser Ser Val		
	190	195	200
30	Glu Cys Leu Glu Leu Arg Asp Thr Asp Leu Asp Thr Phe His Phe Ser		
	205	210	215
	Glu Leu Ser Thr Gly Glu Thr Asn Ser Leu Ile Lys Lys Phe Thr Phe		
	220	225	230
35	Arg Asn Val Lys Ile Thr Asp Glu Ser Leu Phe Gln Val Met Lys Leu		
	235	240	245
	250		
40	Leu Asn Gln Ile Ser Gly Leu Leu Glu Leu Glu Phe Asp Asp Cys Thr		
	255	260	265
	Leu Asn Gly Val Gly Asn Phe Arg Ala Ser Asp Asn Asp Arg Val Ile		
	270	275	280
45	Asp Pro Gly Lys Val Glu Thr Leu Thr Ile Arg Arg Leu His Ile Pro		
	285	290	295
	Arg Phe Tyr Leu Phe Tyr Asp Leu Ser Thr Leu Tyr Ser Leu Thr Glu		
	300	305	310
50	Arg Val Lys Arg Ile Thr Val Glu Asn Ser Lys Val Phe Leu Val Pro		
	315	320	325
	330		
55	Cys Leu Leu Ser Gln His Leu Lys Ser Leu Glu Tyr Leu Asp Leu Ser		
	335	340	345
	Glu Asn Leu Met Val Glu Glu Tyr Leu Lys Asn Ser Ala Cys Glu Asp		
	350	355	360
60	Ala Trp Pro Ser Leu Gln Thr Leu Ile Leu Arg Gln Asn His Leu Ala		
	365	370	375

Ser Leu Glu Lys Thr Gly Glu Thr Leu Leu Thr Leu Lys Asn Leu Thr  
 380 385 390

5 Asn Ile Asp Ile Ser Lys Asn Ser Phe His Ser Met Pro Glu Thr Cys  
 395 400 405 410

Gln Trp Pro Glu Lys Met Lys Tyr Leu Asn Leu Ser Ser Thr Arg Ile  
 10 415 420 425

His Ser Val Thr Gly Cys Ile Pro Lys Thr Leu Glu Ile Leu Asp Val  
 430 435 440

15 Ser Asn Asn Asn Leu Asn Leu Phe Ser Leu Asn Leu Pro Gln Leu Lys  
 445 450 455

Glu Leu Tyr Ile Ser Arg Asn Lys Leu Met Thr Leu Pro Asp Ala Ser  
 460 465 470

20 Leu Leu Pro Met Leu Leu Val Leu Lys Ile Ser Arg Asn Ala Ile Thr  
 475 480 485 490

Thr Phe Ser Lys Glu Gln Leu Asp Ser Phe His Thr Leu Lys Thr Leu  
 25 495 500 505

Glu Ala Gly Gly Asn Asn Phe Ile Cys Ser Cys Glu Phe Leu Ser Phe  
 510 515 520

Thr Gln Glu Gln Gln Ala Leu Ala Lys Val Leu Ile Asp Trp Pro Ala  
 30 525 530 535

Asn Tyr Leu Cys Asp Ser Pro Ser His Val Arg Gly Gln Gln Val Gln  
 540 545 550

35 Asp Val Arg Leu Ser Val Ser Glu Cys His Arg Thr Ala Leu Val Ser  
 555 560 565 570

Gly Met Cys Cys Ala Leu Phe Leu Leu Ile Leu Leu Thr Gly Val Leu  
 40 575 580 585

Cys His Arg Phe His Gly Leu Trp Tyr Met Lys Met Met Trp Ala Trp  
 590 595 600

Leu Gln Ala Lys Arg Lys Pro Arg Lys Ala Pro Ser Arg Asn Ile Cys  
 45 605 610 615

Tyr Asp Ala Phe Val Ser Tyr Ser Glu Arg Asp Ala Tyr Trp Val Glu  
 620 625 630

50 Asn Leu Met Val Gln Glu Leu Glu Asn Phe Asn Pro Pro Phe Lys Leu  
 635 640 645 650

Cys Leu His Lys Arg Asp Phe Ile Pro Gly Lys Trp Ile Ile Asp Asn  
 55 655 660 665

Ile Ile Asp Ser Ile Glu Lys Ser His Lys Thr Val Phe Val Leu Ser  
 670 675 680

60 Glu Asn Phe Val Lys Ser Glu Trp Cys Lys Tyr Glu Leu Asp Phe Ser  
 685 690 695

	His	Phe	Arg	Leu	Phe	Glu	Glu	Asn	Asn	Asp	Ala	Ala	Ile	Leu	Ile	Leu
	700					705								710		
5	Leu	Glu	Pro	Ile	Glu	Lys	Lys	Ala	Ile	Pro	Gln	Arg	Phe	Cys	Lys	Leu
	715				720					725				730		
	Arg	Lys	Ile	Met	Asn	Thr	Lys	Thr	Tyr	Leu	Glu	Trp	Pro	Met	Asp	Glu
					735					740				745		
10	Ala	Gln	Arg	Glu	Gly	Phe	Trp	Val	Asn	Leu	Arg	Ala	Ala	Ile	Lys	Ser
					750				755					760		

(2) INFORMATION FOR SEQ ID NO:5:

15	(i)	SEQUENCE CHARACTERISTICS:
	(A)	LENGTH: 2715 base pairs
	(B)	TYPE: nucleic acid
20	(C)	STRANDEDNESS: single
	(D)	TOPOLOGY: linear

25	(ii)	MOLECULE TYPE: cDNA
	(ix)	FEATURE:
	(A)	NAME/KEY: CDS
	(B)	LOCATION: 1..2712
30	(ix)	FEATURE:
	(A)	NAME/KEY: mat_peptide
	(B)	LOCATION: 64..2712

35	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:5:															
	ATG	AGA	CAG	ACT	TTG	CCT	TGT	ATC	TAC	TTT	TGG	GGG	GGC	CTT	TTG	CCC	48
	Met	Arg	Gln	Thr	Leu	Pro	Cys	Ile	Tyr	Phe	Trp	Gly	Gly	Leu	Leu	Pro	
	-21	-20						-15						-10			
40	TTT	GGG	ATG	CTG	TGT	GCA	TCC	TCC	ACC	ACC	AAG	TGC	ACT	GTT	AGC	CAT	96
	Phe	Gly	Met	Leu	Cys	Ala	Ser	Ser	Thr	Thr	Lys	Cys	Thr	Val	Ser	His	
	-5					1			5					10			
45	GAA	GTT	GCT	GAC	TGC	AGC	CAC	CTG	AAG	TTG	ACT	CAG	GTA	CCC	GAT	GAT	144
	Glu	Val	Ala	Asp	Cys	Ser	His	Leu	Lys	Leu	Thr	Gln	Val	Pro	Asp	Asp	
						15			20				25				
50	CTA	CCC	ACA	AAC	ATA	ACA	GTG	TTG	AAC	CTT	ACC	CAT	AAT	CAA	CTC	AGA	192
	Leu	Pro	Thr	Asn	Ile	Thr	Val	Leu	Asn	Leu	Thr	His	Asn	Gln	Leu	Arg	
						30			35				40				
55	AGA	TTA	CCA	GCC	AAC	TTC	ACA	AGG	TAT	AGC	CAG	CTA	ACT	AGC	TTG		240
	Arg	Leu	Pro	Ala	Ala	Asn	Phe	Thr	Arg	Tyr	Ser	Gln	Leu	Thr	Ser	Leu	
						45			50				55				
60	GAT	GTA	GGA	TTT	AAC	ACC	ATC	TCA	AAA	CTG	GAG	CCA	GAA	TTG	TGC	CAG	288
	Asp	Val	Gly	Phe	Asn	Thr	Ile	Ser	Lys	Leu	Glu	Pro	Glu	Leu	Cys	Gln	
						60			65				70		75		
65	AAA	CTT	CCC	ATG	TTA	AAA	GTT	TTG	AAC	CTC	CAG	CAC	AAT	GAG	CTA	TCT	336
	Lys	Leu	Pro	Met	Leu	Lys	Val	Leu	Asn	Leu	Gln	His	Asn	Glu	Leu	Ser	

	80	85	90	
5	CAA CTT TCT GAT AAA ACC TTT GCC TTC TGC ACG AAT TTG ACT GAA CTC Gln Leu Ser Asp Lys Thr Phe Ala Phe Cys Thr Asn Leu Thr Glu Leu 95 100 105			384
10	CAT CTC ATG TCC AAC TCA ATC CAG AAA ATT AAA AAT AAT CCC TTT GTC His Leu Met Ser Asn Ser Ile Gln Lys Ile Lys Asn Asn Pro Phe Val 110 115 120			432
15	AAG CAG AAG AAT TTA ATC ACA TTA GAT CTG TCT CAT AAT GGC TTG TCA Lys Gln Lys Asn Leu Ile Thr Leu Asp Leu Ser His Asn Gly Leu Ser 125 130 135			480
20	TCT ACA AAA TTA GGA ACT CAG GTT CAG CTG GAA AAT CTC CAA GAG CTT Ser Thr Lys Leu Gly Thr Gln Val Gln Leu Glu Asn Leu Gln Glu Leu 140 145 150 155			528
25	CTA TTA TCA AAC AAT AAA ATT CAA GCG CTA AAA AGT GAA GAA CTG GAT Leu Leu Ser Asn Asn Lys Ile Gln Ala Leu Lys Ser Glu Glu Leu Asp 160 165 170			576
30	ATC TTT GCC AAT TCA TCT TTA AAA AAA TTA GAG TTG TCA TCG AAT CAA Ile Phe Ala Asn Ser Ser Leu Lys Lys Leu Glu Leu Ser Ser Asn Gln 175 180 185			624
35	ATT AAA GAG TTT TCT CCA GGG TGT TTT CAC GCA ATT GGA AGA TTA TTT Ile Lys Glu Phe Ser Pro Gly Cys Phe His Ala Ile Gly Arg Leu Phe 190 195 200			672
40	GGC CTC TTT CTG AAC AAT GTC CAG CTG GGT CCC AGC CTT ACA GAG AAG Gly Leu Phe Leu Asn Asn Val Gln Leu Gly Pro Ser Leu Thr Glu Lys 205 210 215			720
45	CTA TGT TTG GAA TTA GCA AAC ACA AGC ATT CGG AAT CTG TCT CTG AGT Leu Cys Leu Glu Leu Ala Asn Thr Ser Ile Arg Asn Leu Ser Leu Ser 220 225 230 235			768
50	AAC AGC CAG CTG TCC ACC ACC AGC AAT ACA ACT TTC TTG GGA CTA AAG Asn Ser Gln Leu Ser Thr Ser Asn Thr Thr Phe Leu Gly Leu Lys 240 245 250			816
55	TGG ACA AAT CTC ACT ATG CTC GAT CTT TCC TAC AAC AAC TTA AAT GTG Trp Thr Asn Leu Thr Met Leu Asp Leu Ser Tyr Asn Asn Leu Asn Val 255 260 265			864
60	GTT GGT AAC GAT TCC TTT GCT TGG CTT CCA CAA CTA GAA TAT TTC TTC Val Gly Asn Asp Ser Phe Ala Trp Leu Pro Gln Leu Glu Tyr Phe Phe 270 275 280			912
	CTA GAG TAT AAT AAT ATA CAG CAT TTG TTT TCT CAC TCT TTG CAC GGG Leu Glu Tyr Asn Asn Ile Gln His Leu Phe Ser His Ser Leu His Gly 285 290 295			960
	CTT TTC AAT GTG AGG TAC CTG AAT TTG AAA CGG TCT TTT ACT AAA CAA Leu Phe Asn Val Arg Tyr Leu Asn Leu Lys Arg Ser Phe Thr Lys Gln 300 305 310 315			1008
	AGT ATT TCC CTT GCC TCA CTC CCC AAG ATT GAT GAT TTT TCT TTT CAG Ser Ile Ser Leu Ala Ser Leu Pro Lys Ile Asp Asp Phe Ser Phe Gln 320 325 330			1056

	TGG CTA AAA TGT TTG GAG CAC CTT AAC ATG GAA GAT AAT GAT ATT CCA		1104
5	Trp Leu Lys Cys Leu Glu His Leu Asn Met Glu Asp Asn Asp Ile Pro		
	335	340	345
	GGC ATA AAA AGC AAT ATG TTC ACA GGA TTG ATA AAC CTG AAA TAC TTA		1152
	Gly Ile Lys Ser Asn Met Phe Thr Gly Leu Ile Asn Leu Lys Tyr Leu		
	350	355	360
10	AGT CTA TCC AAC TCC TTT ACA AGT TTG CGA ACT TTG ACA AAT GAA ACA		1200
	Ser Leu Ser Asn Ser Phe Thr Ser Leu Arg Thr Leu Thr Asn Glu Thr		
	365	370	375
15	TTT GTA TCA CTT GCT CAT TCT CCC TTA CAC ATA CTC AAC CTA ACC AAG		1248
	Phe Val Ser Leu Ala His Ser Pro Leu His Ile Leu Asn Leu Thr Lys		
	380	385	390
	395		
20	AAT AAA ATC TCA AAA ATA GAG AGT GAT GCT TTC TCT TGG TTG GGC CAC		1296
	Asn Lys Ile Ser Lys Ile Glu Ser Asp Ala Phe Ser Trp Leu Gly His		
	400	405	410
25	CTA GAA GTA CTT GAC CTG GGC CTT AAT GAA ATT GGG CAA GAA CTC ACA		1344
	Leu Glu Val Leu Asp Leu Gly Leu Asn Glu Ile Gly Gln Glu Leu Thr		
	415	420	425
30	GGC CAG GAA TGG AGA GGT CTA GAA AAT ATT TTC GAA ATC TAT CTT TCC		1392
	Gly Gln Glu Trp Arg Gly Leu Glu Asn Ile Phe Glu Ile Tyr Leu Ser		
	430	435	440
35	TAC AAC AAG TAC CTG CAG CTG ACT AGG AAC TCC TTT GCC TTG GTC CCA		1440
	Tyr Asn Lys Tyr Leu Gln Leu Thr Arg Asn Ser Phe Ala Leu Val Pro		
	445	450	455
40	AGC CTT CAA CGA CTG ATG CTC CGA AGG GTG GCC CTT AAA AAT GTG GAT		1488
	Ser Leu Gln Arg Leu Met Leu Arg Arg Val Ala Leu Lys Asn Val Asp		
	460	465	470
	475		
45	AGC TCT CCT TCA CCA TTC CAG CCT CTT CGT AAC TTG ACC ATT CTG GAT		1536
	Ser Ser Pro Ser Pro Phe Gln Pro Leu Arg Asn Leu Thr Ile Leu Asp		
	480	485	490
50	CTA AGC AAC AAC ATA GCC AAC ATA AAT GAT GAC ATG TTG GAG GGT		1584
	Leu Ser Asn Asn Asn Ile Ala Asn Ile Asn Asp Asp Met Leu Glu Gly		
	495	500	505
55	CTT GAG AAA CTA GAA ATT CTC GAT TTG CAG CAT AAC AAC TTA GCA CGG		1632
	Leu Glu Lys Leu Glu Ile Leu Asp Leu Gln His Asn Asn Leu Ala Arg		
	510	515	520
60	CTC TGG AAA CAC GCA AAC CCT GGT GGT CCC ATT TAT TTC CTA AAG GGT		1680
	Leu Trp Lys His Ala Asn Pro Gly Gly Pro Ile Tyr Phe Leu Lys Gly		
	525	530	535
65	CTG TCT CAC CTC CAC ATC CTT AAC TTG GAG TCC AAC GGC TTT GAC GAG		1728
	Leu Ser His Leu His Ile Leu Asn Leu Glu Ser Asn Gly Phe Asp Glu		
	540	545	550
	555		
70	ATC CCA GTT GAG GTC TTC AAG GAT TTA TTT GAA CTA AAG ATC ATC GAT		1776
	Ile Pro Val Glu Val Phe Lys Asp Leu Phe Glu Leu Lys Ile Ile Asp		
	560	565	570

	TTA GGA TTG AAT AAT TTA AAC ACA CTT CCA GCA TCT GTC TTT AAT AAT Leu Gly Leu Asn Asn Leu Asn Thr Leu Pro Ala Ser Val Phe Asn Asn 575 580 585	1824
5	CAG GTG TCT CTA AAG TCA TTG AAC CTT CAG AAG AAT CTC ATA ACA TCC Gln Val Ser Leu Lys Ser Leu Asn Leu Gln Lys Asn Leu Ile Thr Ser 590 595 600	1872
10	GTT GAG AAG AAG GTT TTC GGG CCA GCT TTC AGG AAC CTG ACT GAG TTA Val Glu Lys Lys Val Phe Gly Pro Ala Phe Arg Asn Leu Thr Glu Leu 605 610 615	1920
15	GAT ATG CGC TTT AAT CCC TTT GAT TGC ACG TGT GAA AGT ATT GCC TGG Asp Met Arg Phe Asn Pro Phe Asp Cys Thr Cys Glu Ser Ile Ala Trp 620 625 630 635	1968
20	TTT GTT AAT TGG ATT AAC GAG ACC CAT ACC AAC ATC CCT GAG CTG TCA Phe Val Asn Trp Ile Asn Glu Thr His Thr Asn Ile Pro Glu Leu Ser 640 645 650	2016
25	AGC CAC TAC CTT TGC AAC ACT CCA CCT CAC TAT CAT GGG TTC CCA GTG Ser His Tyr Leu Cys Asn Thr Pro Pro His Tyr His Gly Phe Pro Val 655 660 665	2064
30	AGA CTT TTT GAT ACA TCA TCT TGC AAA GAC AGT GCC CCC TTT GAA CTC Arg Leu Phe Asp Thr Ser Ser Cys Lys Asp Ser Ala Pro Phe Glu Leu 670 675 680	2112
35	TTT TTC ATG ATC AAT ACC AGT ATC CTG TTG ATT TTT ATC TTT ATT GTA Phe Phe Met Ile Asn Thr Ser Ile Leu Leu Ile Phe Ile Phe Ile Val 685 690 695	2160
40	CTT CTC ATC CAC TTT GAG GGC TGG AGG ATA TCT TTT TAT TGG AAT GTT Leu Leu Ile His Phe Glu Gly Trp Arg Ile Ser Phe Tyr Trp Asn Val 700 705 710 715	2208
45	TCA GTA CAT CGA GTT CTT GGT TTC AAA GAA ATA GAC AGA CAG ACA GAA Ser Val His Arg Val Leu Gly Phe Lys Glu Ile Asp Arg Gln Thr Glu 720 725 730	2256
50	CAG TTT GAA TAT GCA GCA TAT ATA ATT CAT GCC TAT AAA GAT AAG GAT Gln Phe Glu Tyr Ala Ala Tyr Ile Ile His Ala Tyr Lys Asp Lys Asp 735 740 745	2304
55	TGG GTC TGG GAA CAT TTC TCT TCA ATG GAA AAG GAA GAC CAA TCT CTC Trp Val Trp Glu His Phe Ser Ser Met Glu Lys Glu Asp Gln Ser Leu 750 755 760	2352
60	AAA TTT TGT CTG GAA GAA AGG GAC TTT GAG GCG GGT GTT TTT GAA CTA Lys Phe Cys Leu Glu Glu Arg Asp Phe Glu Ala Gly Val Phe Glu Leu 765 770 775	2400
	GAA GCA ATT GTT AAC AGC ATC AAA AGA AGC AGA AAA ATT ATT TTT GTT Glu Ala Ile Val Asn Ser Ile Lys Arg Ser Arg Lys Ile Ile Phe Val	2448
	780 785 790 795	
	ATA ACA CAC CAT CTA TTA AAA GAC CCA TTA TGC AAA AGA TTC AAG GTA Ile Thr His His Leu Leu Lys Asp Pro Leu Cys Lys Arg Phe Lys Val 800 805 810	2496
	CAT CAT GCA GTT CAA CAA GCT ATT GAA CAA AAT CTG GAT TCC ATT ATA	2544

	His His Ala Val Gln Gln Ala Ile Glu Gln Asn Leu Asp Ser Ile Ile		
	815	820	825
5	TTG GTT TTC CTT GAG GAG ATT CCA GAT TAT AAA CTG AAC CAT GCA CTC		2592
	Leu Val Phe Leu Glu Glu Ile Pro Asp Tyr Lys Leu Asn His Ala Leu		
	830	835	840
10	TGT TTG CGA AGA GGA ATG TTT AAA TCT CAC TGC ATC TTG AAC TGG CCA		2640
	Cys Leu Arg Arg Gly Met Phe Lys Ser His Cys Ile Leu Asn Trp Pro		
	845	850	855
15	GTT CAG AAA GAA CGG ATA GGT GCC TTT CGT CAT AAA TTG CAA GTA GCA		2688
	Val Gln Lys Glu Arg Ile Gly Ala Phe Arg His Lys Leu Gln Val Ala		
	860	865	870
	875		
20	CTT GGA TCC AAA AAC TCT GTA CAT TAA		2715
	Leu Gly Ser Lys Asn Ser Val His		
	880		
25	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 904 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
30	(ii) MOLECULE TYPE: protein		
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:		
	Met Arg Gln Thr Leu Pro Cys Ile Tyr Phe Trp Gly Gly Leu Leu Pro		
	-21 -20	-15	-10
40	Phe Gly Met Leu Cys Ala Ser Ser Thr Thr Lys Cys Thr Val Ser His		
	-5	1	5
	10		
	Glu Val Ala Asp Cys Ser His Leu Lys Leu Thr Gln Val Pro Asp Asp		
	15	20	25
45	Leu Pro Thr Asn Ile Thr Val Leu Asn Leu Thr His Asn Gln Leu Arg		
	30	35	40
	Arg Leu Pro Ala Ala Asn Phe Thr Arg Tyr Ser Gln Leu Thr Ser Leu		
	45	50	55
	Asp Val Gly Phe Asn Thr Ile Ser Lys Leu Glu Pro Glu Leu Cys Gln		
	60	65	70
50	Lys Leu Pro Met Leu Lys Val Leu Asn Leu Gln His Asn Glu Leu Ser		
	80	85	90
55	Gln Leu Ser Asp Lys Thr Phe Ala Phe Cys Thr Asn Leu Thr Glu Leu		
	95	100	105
	His Leu Met Ser Asn Ser Ile Gln Lys Ile Lys Asn Asn Pro Phe Val		
	110	115	120
60	Lys Gln Lys Asn Leu Ile Thr Leu Asp Leu Ser His Asn Gly Leu Ser		
	125	130	135

Ser Thr Lys Leu Gly Thr Gln Val Gln Leu Glu Asn Leu Gln Glu Leu  
 140 145 150 155

5 Leu Leu Ser Asn Asn Lys Ile Gln Ala Leu Lys Ser Glu Glu Leu Asp  
 160 165 170

Ile Phe Ala Asn Ser Ser Leu Lys Lys Leu Glu Leu Ser Ser Asn Gln  
 175 180 185

10 Ile Lys Glu Phe Ser Pro Gly Cys Phe His Ala Ile Gly Arg Leu Phe  
 190 195 200

Gly Leu Phe Leu Asn Asn Val Gln Leu Gly Pro Ser Leu Thr Glu Lys  
 205 210 215

15 Leu Cys Leu Glu Leu Ala Asn Thr Ser Ile Arg Asn Leu Ser Leu Ser  
 220 225 230 235

Asn Ser Gln Leu Ser Thr Thr Ser Asn Thr Thr Phe Leu Gly Leu Lys  
 240 245 250

Trp Thr Asn Leu Thr Met Leu Asp Leu Ser Tyr Asn Asn Leu Asn Val  
 255 260 265

25 Val Gly Asn Asp Ser Phe Ala Trp Leu Pro Gln Leu Glu Tyr Phe Phe  
 270 275 280

Leu Glu Tyr Asn Asn Ile Gln His Leu Phe Ser His Ser Leu His Gly  
 285 290 295

30 Leu Phe Asn Val Arg Tyr Leu Asn Leu Lys Arg Ser Phe Thr Lys Gln  
 300 305 310 315

Ser Ile Ser Leu Ala Ser Leu Pro Lys Ile Asp Asp Phe Ser Phe Gln  
 320 325 330

Trp Leu Lys Cys Leu Glu His Leu Asn Met Glu Asp Asn Asp Ile Pro  
 335 340 345

40 Gly Ile Lys Ser Asn Met Phe Thr Gly Leu Ile Asn Leu Lys Tyr Leu  
 350 355 360

Ser Leu Ser Asn Ser Phe Thr Ser Leu Arg Thr Leu Thr Asn Glu Thr  
 365 370 375

45 Phe Val Ser Leu Ala His Ser Pro Leu His Ile Leu Asn Leu Thr Lys  
 380 385 390 395

Asn Lys Ile Ser Lys Ile Glu Ser Asp Ala Phe Ser Trp Leu Gly His  
 400 405 410

Leu Glu Val Leu Asp Leu Gly Leu Asn Glu Ile Gly Gln Glu Leu Thr  
 415 420 425

55 Gly Gln Glu Trp Arg Gly Leu Glu Asn Ile Phe Glu Ile Tyr Leu Ser  
 430 435 440

Tyr Asn Lys Tyr Leu Gln Leu Thr Arg Asn Ser Phe Ala Leu Val Pro  
 445 450 455

60 Ser Leu Gln Arg Leu Met Leu Arg Arg Val Ala Leu Lys Asn Val Asp

	460	465	470	475
	Ser Ser Pro Ser Pro Phe Gln Pro Leu Arg Asn Leu Thr Ile Leu Asp			
	480	485	490	
5	Leu Ser Asn Asn Asn Ile Ala Asn Ile Asn Asp Asp Met Leu Glu Gly			
	495	500	505	
10	Leu Glu Lys Leu Glu Ile Leu Asp Leu Gln His Asn Asn Leu Ala Arg			
	510	515	520	
	Leu Trp Lys His Ala Asn Pro Gly Gly Pro Ile Tyr Phe Leu Lys Gly			
	525	530	535	
15	Leu Ser His Leu His Ile Leu Asn Leu Glu Ser Asn Gly Phe Asp Glu			
	540	545	550	555
	Ile Pro Val Glu Val Phe Lys Asp Leu Phe Glu Leu Lys Ile Ile Asp			
	560	565	570	
20	Leu Gly Leu Asn Asn Leu Asn Thr Leu Pro Ala Ser Val Phe Asn Asn			
	575	580	585	
	Gln Val Ser Leu Lys Ser Leu Asn Leu Gln Lys Asn Leu Ile Thr Ser			
25	590	595	600	
	Val Glu Lys Lys Val Phe Gly Pro Ala Phe Arg Asn Leu Thr Glu Leu			
	605	610	615	
30	Asp Met Arg Phe Asn Pro Phe Asp Cys Thr Cys Glu Ser Ile Ala Trp			
	620	625	630	635
	Phe Val Asn Trp Ile Asn Glu Thr His Thr Asn Ile Pro Glu Leu Ser			
	640	645	650	
35	Ser His Tyr Leu Cys Asn Thr Pro Pro His Tyr His Gly Phe Pro Val			
	655	660	665	
	Arg Leu Phe Asp Thr Ser Ser Cys Lys Asp Ser Ala Pro Phe Glu Leu			
40	670	675	680	
	Phe Phe Met Ile Asn Thr Ser Ile Leu Leu Ile Phe Ile Phe Ile Val			
	685	690	695	
45	Leu Leu Ile His Phe Glu Gly Trp Arg Ile Ser Phe Tyr Trp Asn Val			
	700	705	710	715
	Ser Val His Arg Val Leu Gly Phe Lys Glu Ile Asp Arg Gln Thr Glu			
	720	725	730	
50	Gln Phe Glu Tyr Ala Ala Tyr Ile Ile His Ala Tyr Lys Asp Lys Asp			
	735	740	745	
	Trp Val Trp Glu His Phe Ser Ser Met Glu Lys Glu Asp Gln Ser Leu			
55	750	755	760	
	Lys Phe Cys Leu Glu Glu Arg Asp Phe Glu Ala Gly Val Phe Glu Leu			
	765	770	775	
60	Glu Ala Ile Val Asn Ser Ile Lys Arg Ser Arg Lys Ile Ile Phe Val			
	780	785	790	795

Ile Thr His His Leu Leu Lys Asp Pro Leu Cys Lys Arg Phe Lys Val  
 800 805 810

5 His His Ala Val Gln Gln Ala Ile Glu Gln Asn Leu Asp Ser Ile Ile  
 815 820 825

Leu Val Phe Leu Glu Glu Ile Pro Asp Tyr Lys Leu Asn His Ala Leu  
 830 835 840

10 Cys Leu Arg Arg Gly Met Phe Lys Ser His Cys Ile Leu Asn Trp Pro  
 845 850 855

15 Val Gln Lys Glu Arg Ile Gly Ala Phe Arg His Lys Leu Gln Val Ala  
 860 865 870 875

Leu Gly Ser Lys Asn Ser Val His  
 880

## 20 (2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2400 base pairs
- (B) TYPE: nucleic acid
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30 (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2397

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATG GAG CTG AAT TTC TAC AAA ATC CCC GAC AAC CTC CCC TTC TCA ACC	48
Met Glu Leu Asn Phe Tyr Lys Ile Pro Asp Asn Leu Pro Phe Ser Thr	
40 1 5 10 15	
AAG AAC CTG GAC CTG AGC TTT AAT CCC CTG AGG CAT TTA GGC AGC TAT	96
Lys Asn Leu Asp Leu Ser Phe Asn Pro Leu Arg His Leu Gly Ser Tyr	
45 20 25 30	
AGC TTC TTC AGT TTC CCA GAA CTG CAG GTG CTG GAT TTA TCC AGG TGT	144
Ser Phe Phe Ser Phe Pro Glu Leu Gln Val Leu Asp Leu Ser Arg Cys	
35 40 45	
50 GAA ATC CAG ACA ATT GAA GAT GGG GCA TAT CAG AGC CTA AGC CAC CTC	192
Glu Ile Gln Thr Ile Glu Asp Gly Ala Tyr Gln Ser Leu Ser His Leu	
55 50 55 60	
TCT ACC TTA ATA TTG ACA GGA AAC CCC ATC CAG AGT TTA GCC CTG GGA	240
Ser Thr Leu Ile Leu Thr Gly Asn Pro Ile Gln Ser Leu Ala Leu Gly	
60 65 70 75 80	
GCC TTT TCT GGA CTA TCA AGT TTA CAG AAG CTG GTG GCT GTG GAG ACA	288
Ala Phe Ser Gly Leu Ser Ser Leu Gln Lys Leu Val Ala Val Glu Thr	
65 85 90 95	

	AAT CTA GCA TCT CTA GAG AAC TTC CCC ATT GGA CAT CTC AAA ACT TTG Asn Leu Ala Ser Leu Glu Asn Phe Pro Ile Gly His Leu Lys Thr Leu 100 105 110	336
5	AAA GAA CTT AAT GTG GCT CAC AAT CTT ATC CAA TCT TTC AAA TTA CCT Lys Glu Leu Asn Val Ala His Asn Leu Ile Gln Ser Phe Lys Leu Pro 115 120 125	384
10	GAG TAT TTT TCT AAT CTG ACC AAT CTA GAG CAC TTG GAC CTT TCC AGC Glu Tyr Phe Ser Asn Leu Thr Asn Leu Glu His Leu Asp Leu Ser Ser. 130 135 140	432
15	AAC AAG ATT CAA AGT ATT TAT TGC ACA GAC TTG CGG GTT CTA CAT CAA Asn Lys Ile Gln Ser Ile Tyr Cys Thr Asp Leu Arg Val Leu His Gln 145 150 155 160	480
20	ATG CCC CTA CTC AAT CTC TCT TTA GAC CTG TCC CTG AAC CCT ATG AAC Met Pro Leu Leu Asn Leu Ser Leu Asp Leu Ser Leu Asn Pro Met Asn 165 170 175	528
25	TTT ATC CAA CCA GGT GCA TTT AAA GAA ATT AGG CTT CAT AAG CTG ACT Phe Ile Gln Pro Gly Ala Phe Lys Glu Ile Arg Leu His Lys Leu Thr 180 185 190	576
30	TTA AGA AAT AAT TTT GAT AGT TTA AAT GTA ATG AAA ACT TGT ATT CAA Leu Arg Asn Asn Phe Asp Ser Leu Asn Val Met Lys Thr Cys Ile Gln 195 200 205	624
35	GGT CTG GCT GGT TTA GAA GTC CAT CGT TTG GTT CTG GGA GAA TTT AGA Gly Leu Ala Gly Leu Glu Val His Arg Leu Val Leu Gly Glu Phe Arg 210 215 220	672
40	AAT GAA GGA AAC TTG GAA AAG TTT GAC AAA TCT GCT CTA GAG GGC CTG Asn Glu Gly Asn Leu Glu Lys Phe Asp Lys Ser Ala Leu Glu Gly Leu 225 230 235 240	720
45	TGC AAT TTG ACC ATT GAA GAA TTC CGA TTA GCA TAC TTA GAC TAC TAC Cys Asn Leu Thr Ile Glu Glu Phe Arg Leu Ala Tyr Leu Asp Tyr Tyr 245 250 255	768
50	CTC GAT GAT ATT ATT GAC TTA TTT AAT TGT TTG ACA AAT GTT TCT TCA Leu Asp Asp Ile Ile Asp Leu Phe Asn Cys Leu Thr Asn Val Ser Ser 260 265 270	816
55	TTT TCC CTG GTG AGT GTG ACT ATT GAA AGG GTA AAA GAC TTT TCT TAT Phe Ser Leu Val Ser Val Thr Ile Glu Arg Val Lys Asp Phe Ser Tyr 275 280 285	864
60	AAT TTC GGA TGG CAA CAT TTA GAA TTA GTT AAC TGT AAA TTT GGA CAG Asn Phe Gly Trp Gln His Leu Glu Leu Val Asn Cys Lys Phe Gly Gln 290 295 300	912
	TTT CCC ACA TTG AAA CTC AAA TCT CTC AAA AGG CTT ACT TTC ACT TCC Phe Pro Thr Leu Lys Leu Lys Ser Leu Lys Arg Leu Thr Phe Thr Ser 305 310 315 320	960
	AAC AAA GGT GGG AAT GCT TTT TCA GAA GTT GAT CTA CCA AGC CTT GAG Asn Lys Gly Gly Asn Ala Phe Ser Glu Val Asp Leu Pro Ser Leu Glu 325 330 335	1008
	TTT CTA GAT CTC AGT AGA AAT GGC TTG AGT TTC AAA GGT TGC TGT TCT	1056

	Phe Leu Asp Leu Ser Arg Asn Gly Leu Ser Phe Lys Gly Cys Cys Ser		
	340 345 350		
5	CAA AGT GAT TTT GGG ACA ACC AGC CTA AAG TAT TTA GAT CTG AGC TTC Gln Ser Asp Phe Gly Thr Thr Ser Leu Lys Tyr Leu Asp Leu Ser Phe	355 360 365	1104
10	AAT GGT GTT ATT ACC ATG AGT TCA AAC TTC TTG GGC TTA GAA CAA CTA Asn Gly Val Ile Thr Met Ser Ser Asn Phe Leu Gly Leu Glu Gln Leu	370 375 380	1152
15	GAA CAT CTG GAT TTC CAG CAT TCC AAT TTG AAA CAA ATG AGT GAG TTT Glu His Leu Asp Phe Gln His Ser Asn Leu Lys Gln Met Ser Glu Phe	385 390 395 400	1200
20	TCA GTA TTC CTA TCA CTC AGA AAC CTC ATT TAC CTT GAC ATT TCT CAT Ser Val Phe Leu Ser Leu Arg Asn Leu Ile Tyr Leu Asp Ile Ser His	405 410 415	1248
25	ACT CAC ACC AGA GTT GCT TTC AAT GGC ATC TTC AAT GGC TTG TCC AGT Thr His Thr Arg Val Ala Phe Asn Gly Ile Phe Asn Gly Leu Ser Ser	420 425 430	1296
30	CTC GAA GTC TTG AAA ATG GCT GGC AAT TCT TTC CAG GAA AAC TTC CTT Leu Glu Val Leu Lys Met Ala Gly Asn Ser Phe Gln Glu Asn Phe Leu	435 440 445	1344
35	CCA GAT ATC TTC ACA GAG CTG AGA AAC TTG ACC TTC CTG GAC CTC TCT Pro Asp Ile Phe Thr Glu Leu Arg Asn Leu Thr Phe Leu Asp Leu Ser	450 455 460	1392
40	CAG TGT CAA CTG GAG CAG TTG TCT CCA ACA GCA TTT AAC TCA CTC TCC Gln Cys Gln Leu Glu Gln Leu Ser Pro Thr Ala Phe Asn Ser Leu Ser	465 470 475 480	1440
45	AGT CTT CAG GTA CTA AAT ATG AGC CAC AAC AAC TTC TTT TCA TTG GAT Ser Leu Gln Val Leu Asn Met Ser His Asn Asn Phe Phe Ser Leu Asp	485 490 495	1488
50	ACG TTT CCT TAT AAG TGT CTG AAC TCC CTC CAG GTT CTT GAT TAC AGT Thr Phe Pro Tyr Lys Cys Leu Asn Ser Leu Gln Val Leu Asp Tyr Ser	500 505 510	1536
55	CTC AAT CAC ATA ATG ACT TCC AAA AAA CAG GAA CTA CAG CAT TTT CCA Leu Asn His Ile Met Thr Ser Lys Lys Gln Glu Leu Gln His Phe Pro	515 520 525	1584
60	AGT AGT CTA GCT TTC TTA AAT CTT ACT CAG AAT GAC TTT GCT TGT ACT Ser Ser Leu Ala Phe Leu Asn Leu Thr Gln Asn Asp Phe Ala Cys Thr	530 535 540	1632
	TGT GAA CAC CAG AGT TTC CTG CAA TGG ATC AAG GAC CAG AGG CAG CTC Cys Glu His Gln Ser Phe Leu Gln Trp Ile Lys Asp Gln Arg Gln Leu	545 550 555 560	1680
	TTG GTG GAA GTT GAA CGA ATG GAA TGT GCA ACA CCT TCA GAT AAG CAG Leu Val Glu Val Glu Arg Met Glu Cys Ala Thr Pro Ser Asp Lys Gln	565 570 575	1728
	GGC ATG CCT GTG CTG AGT TTG AAT ATC ACC TGT CAG ATG AAT AAG ACC Gly Met Pro Val Leu Ser Leu Asn Ile Thr Cys Gln Met Asn Lys Thr		1776

	580	585	590	
5	ATC ATT GGT GTG TCG GTC CTC AGT GTG CTT GTA GTA TCT GTT GTA GCA Ile Ile Gly Val Ser Val Leu Ser Val Leu Val Val Ser Val Val Ala 595 600 605			1824
10	GTT CTG GTC TAT AAG TTC TAT TTT CAC CTG ATG CTT CTT GCT GGC TGC Val Leu Val Tyr Lys Phe Tyr Phe His Leu Met Leu Leu Ala Gly Cys 610 615 620			1872
15	ATA AAG TAT GGT AGA GGT GAA AAC ATC TAT GAT GCC TTT GTT ATC TAC Ile Lys Tyr Gly Arg Gly Glu Asn Ile Tyr Asp Ala Phe Val Ile Tyr 625 630 635 640			1920
20	TCA AGC CAG GAT GAG GAC TGG GTA AGG AAT GAG CTA GTA AAG AAT TTA Ser Ser Gln Asp Glu Asp Trp Val Arg Asn Glu Leu Val Lys Asn Leu 645 650 655			1968
25	GAA GAA GGG GTG CCT CCA TTT CAG CTC TGC CTT CAC TAC AGA GAC TTT Glu Glu Gly Val Pro Pro Phe Gln Leu Cys Leu His Tyr Arg Asp Phe 660 665 670			2016
30	ATT CCC GGT GTG GCC ATT GCT GCC AAC ATC ATC CAT GAA GGT TTC CAT Ile Pro Gly Val Ala Ile Ala Ala Asn Ile Ile His Glu Gly Phe His 675 680 685			2064
35	AAA AGC CGA AAG GTG ATT GTT GTG GTG TCC CAG CAC TTC ATC CAG AGC Lys Ser Arg Lys Val Ile Val Val Val Ser Gln His Phe Ile Gln Ser 690 695 700			2112
40	CGC TGG TGT ATC TTT GAA TAT GAG ATT GCT CAG ACC TGG CAG TTT CTG Arg Trp Cys Ile Phe Glu Tyr Glu Ile Ala Gln Thr Trp Gln Phe Leu 705 710 715 720			2160
45	AGC AGT CGT GCT GGT ATC ATC TTC ATT GTC CTG CAG AAG GTG GAG AAG Ser Ser Arg Ala Gly Ile Ile Phe Ile Val Leu Gln Lys Val Glu Lys 725 730 735			2208
50	ACC CTG CTC AGG CAG CAG GTG GAG CTG TAC CGC CTT CTC AGC AGG AAC Thr Leu Leu Arg Gln Gln Val Glu Leu Tyr Arg Leu Leu Ser Arg Asn 740 745 750			2256
55	ACT TAC CTG GAG TGG GAG GAC AGT GTC CTG GGG CGG CAC ATC TTC TGG Thr Tyr Leu Glu Trp Glu Asp Ser Val Leu Gly Arg His Ile Phe Trp 755 760 765			2304
60	AGA CGA CTC AGA AAA GCC CTG CTG GAT GGT AAA TCA TGG AAT CCA GAA Arg Arg Leu Arg Lys Ala Leu Leu Asp Gly Lys Ser Trp Asn Pro Glu 770 775 780			2352
	GGA ACA GTG GGT ACA GGA TGC AAT TGG CAG GAA GCA ACA TCT ATC Gly Thr Val Gly Thr Gly Cys Asn Trp Gln Glu Ala Thr Ser Ile 785 790 795			2397
	TGA			2400

## (2) INFORMATION FOR SEQ ID NO:8:

- 60 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 799 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Glu	Leu	Asn	Phe	Tyr	Lys	Ile	Pro	Asp	Asn	Leu	Pro	Phe	Ser	Thr	
1				5					10				15			
10	Lys	Asn	Leu	Asp	Leu	Ser	Phe	Asn	Pro	Leu	Arg	His	Leu	Gly	Ser	Tyr
				20					25				30			
15	Ser	Phe	Phe	Ser	Phe	Pro	Glu	Leu	Gln	Val	Leu	Asp	Leu	Ser	Arg	Cys
				35				40				45				
20	Glu	Ile	Gln	Thr	Ile	Glu	Asp	Gly	Ala	Tyr	Gln	Ser	Leu	Ser	His	Leu
				50				55				60				
25	Ser	Thr	Leu	Ile	Leu	Thr	Gly	Asn	Pro	Ile	Gln	Ser	Leu	Ala	Leu	Gly
				65			70				75				80	
30	Ala	Phe	Ser	Gly	Leu	Ser	Ser	Leu	Gln	Lys	Leu	Val	Ala	Val	Glu	Thr
				85					90				95			
35	Asn	Leu	Ala	Ser	Leu	Glu	Asn	Phe	Pro	Ile	Gly	His	Leu	Lys	Thr	Leu
				100					105				110			
40	Lys	Glu	Leu	Asn	Val	Ala	His	Asn	Leu	Ile	Gln	Ser	Phe	Lys	Leu	Pro
				115				120				125				
45	Glu	Tyr	Phe	Ser	Asn	Leu	Thr	Asn	Leu	Glu	His	Leu	Asp	Leu	Ser	Ser
				130				135				140				
50	Asn	Lys	Ile	Gln	Ser	Ile	Tyr	Cys	Thr	Asp	Leu	Arg	Val	Leu	His	Gln
				145			150				155				160	
55	Met	Pro	Leu	Leu	Asn	Leu	Ser	Leu	Asp	Leu	Ser	Leu	Asn	Pro	Met	Asn
					165					170				175		
60	Phe	Ile	Gln	Pro	Gly	Ala	Phe	Lys	Glu	Ile	Arg	Leu	His	Lys	Leu	Thr
					180				185				190			
65	Leu	Arg	Asn	Asn	Phe	Asp	Ser	Leu	Asn	Val	Met	Lys	Thr	Cys	Ile	Gln
					195				200				205			
70	Gly	Leu	Ala	Gly	Leu	Glu	Val	His	Arg	Leu	Val	Leu	Gly	Glu	Phe	Arg
					210			215				220				
75	Asn	Glu	Gly	Asn	Leu	Glu	Lys	Phe	Asp	Lys	Ser	Ala	Leu	Glu	Gly	Leu
					225			230				235				240
80	Cys	Asn	Leu	Thr	Ile	Glu	Glu	Phe	Arg	Leu	Ala	Tyr	Leu	Asp	Tyr	Tyr
					245				250				255			
85	Leu	Asp	Asp	Ile	Ile	Asp	Leu	Phe	Asn	Cys	Leu	Thr	Asn	Val	Ser	Ser
					260				265				270			
90	Phe	Ser	Leu	Val	Ser	Val	Thr	Ile	Glu	Arg	Val	Lys	Asp	Phe	Ser	Tyr
					275				280				285			

	Asn Phe Gly Trp Gln His Leu Glu Leu Val Asn Cys Lys Phe Gly Gln			
	290	295	300	
5	Phe Pro Thr Leu Lys Leu Lys Ser Leu Lys Arg Leu Thr Phe Thr Ser			
	305	310	315	320
	Asn Lys Gly Gly Asn Ala Phe Ser Glu Val Asp Leu Pro Ser Leu Glu			
	325	330	335	
10	Phe Leu Asp Leu Ser Arg Asn Gly Leu Ser Phe Lys Gly Cys Cys Ser.			
	340	345	350	
	Gln Ser Asp Phe Gly Thr Thr Ser Leu Lys Tyr Leu Asp Leu Ser Phe			
	355	360	365	
15	Asn Gly Val Ile Thr Met Ser Ser Asn Phe Leu Gly Leu Glu Gln Leu			
	370	375	380	
20	Glu His Leu Asp Phe Gln His Ser Asn Leu Lys Gln Met Ser Glu Phe			
	385	390	395	400
	Ser Val Phe Leu Ser Leu Arg Asn Leu Ile Tyr Leu Asp Ile Ser His			
	405	410	415	
25	Thr His Thr Arg Val Ala Phe Asn Gly Ile Phe Asn Gly Leu Ser Ser			
	420	425	430	
	Leu Glu Val Leu Lys Met Ala Gly Asn Ser Phe Gln Glu Asn Phe Leu			
	435	440	445	
30	Pro Asp Ile Phe Thr Glu Leu Arg Asn Leu Thr Phe Leu Asp Leu Ser			
	450	455	460	
35	Gln Cys Gln Leu Glu Gln Leu Ser Pro Thr Ala Phe Asn Ser Leu Ser			
	465	470	475	480
	Ser Leu Gln Val Leu Asn Met Ser His Asn Asn Phe Phe Ser Leu Asp			
	485	490	495	
40	Thr Phe Pro Tyr Lys Cys Leu Asn Ser Leu Gln Val Leu Asp Tyr Ser			
	500	505	510	
	Leu Asn His Ile Met Thr Ser Lys Lys Gln Glu Leu Gln His Phe Pro			
	515	520	525	
45	Ser Ser Leu Ala Phe Leu Asn Leu Thr Gln Asn Asp Phe Ala Cys Thr			
	530	535	540	
50	Cys Glu His Gln Ser Phe Leu Gln Trp Ile Lys Asp Gln Arg Gln Leu			
	545	550	555	560
	Leu Val Glu Val Glu Arg Met Glu Cys Ala Thr Pro Ser Asp Lys Gln			
	565	570	575	
55	Gly Met Pro Val Leu Ser Leu Asn Ile Thr Cys Gln Met Asn Lys Thr			
	580	585	590	
	Ile Ile Gly Val Ser Val Leu Ser Val Leu Val Val Ser Val Val Ala			
	595	600	605	
60	Val Leu Val Tyr Lys Phe Tyr Phe His Leu Met Leu Leu Ala Gly Cys			

	610	615	620
	Ile Lys Tyr Gly Arg Gly Glu Asn Ile Tyr Asp Ala Phe Val Ile Tyr		
5	625	630	635
	Ser Ser Gln Asp Glu Asp Trp Val Arg Asn Glu Leu Val Lys Asn Leu		
	645	650	655
10	Glu Glu Gly Val Pro Pro Phe Gln Leu Cys Leu His Tyr Arg Asp Phe		
	660	665	670
	Ile Pro Gly Val Ala Ile Ala Ala Asn Ile Ile His Glu Gly Phe His		
	675	680	685
15	Lys Ser Arg Lys Val Ile Val Val Val Ser Gln His Phe Ile Gln Ser		
	690	695	700
	Arg Trp Cys Ile Phe Glu Tyr Glu Ile Ala Gln Thr Trp Gln Phe Leu		
20	705	710	715
	Ser Ser Arg Ala Gly Ile Ile Phe Ile Val Leu Gln Lys Val Glu Lys		
	725	730	735
25	Thr Leu Leu Arg Gln Gln Val Glu Leu Tyr Arg Leu Leu Ser Arg Asn		
	740	745	750
	Thr Tyr Leu Glu Trp Glu Asp Ser Val Leu Gly Arg His Ile Phe Trp		
	755	760	765
30	Arg Arg Leu Arg Lys Ala Leu Leu Asp Gly Lys Ser Trp Asn Pro Glu		
	770	775	780
	Gly Thr Val Gly Thr Gly Cys Asn Trp Gln Glu Ala Thr Ser Ile		
	785	790	795
35	(2) INFORMATION FOR SEQ ID NO:9:		
	(i) SEQUENCE CHARACTERISTICS:		
40	(A) LENGTH: 1275 base pairs		
	(B) TYPE: nucleic acid		
	(C) STRANDEDNESS: single		
	(D) TOPOLOGY: linear		
45	(ii) MOLECULE TYPE: cDNA		
	(ix) FEATURE:		
50	(A) NAME/KEY: CDS		
	(B) LOCATION: 1..1095		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:		
55	TGT TGG GAT GTT TTT GAG GGA CTT TCT CAT CTT CAA GTT CTG TAT TTG		48
	Cys Trp Asp Val Phe Glu Gly Leu Ser His Leu Gln Val Leu Tyr Leu		
	1	5	10
	15		
60	AAT CAT AAC TAT CTT AAT TCC CTT CCA CCA GGA GTA TTT AGC CAT CTG		96
	Asn His Asn Tyr Leu Asn Ser Leu Pro Pro Gly Val Phe Ser His Leu		
	20	25	30

	ACT GCA TTA AGG GGA CTA AGC CTC AAC TCC AAC AGG CTG ACA GTT CTT	144
	Thr Ala Leu Arg Gly Leu Ser Leu Asn Ser Asn Arg Leu Thr Val Leu	
	35 40 45	
5	TCT CAC AAT GAT TTA CCT GCT AAT TTA GAG ATC CTG GAC ATA TCC AGG	192
	Ser His Asn Asp Leu Pro Ala Asn Leu Glu Ile Leu Asp Ile Ser Arg	
	50 55 60	
10	AAC CAG CTC CTA GCT CCT AAT CCT GAT GTA TTT GTA TCA CTT AGT GTC	240
	Asn Gln Leu Leu Ala Pro Asn Pro Asp Val Phe Val Ser Leu Ser Val	
	65 70 75 80	
15	TTG GAT ATA ACT CAT AAC AAG TTC ATT TGT GAA TGT GAA CTT AGC ACT	288
	Leu Asp Ile Thr His Asn Lys Phe Ile Cys Glu Cys Glu Leu Ser Thr	
	85 90 95	
20	TTT ATC AAT TGG CTT AAT CAC ACC AAT GTC ACT ATA GCT GGG CCT CCT	336
	Phe Ile Asn Trp Leu Asn His Thr Asn Val Thr Ile Ala Gly Pro Pro	
	100 105 110	
25	GCA GAC ATA TAT TGT GTG TAC CCT GAC TCG TTC TCT GGG GTT TCC CTC	384
	Ala Asp Ile Tyr Cys Val Tyr Pro Asp Ser Phe Ser Gly Val Ser Leu	
	115 120 125	
30	TTC TCT CTT TCC ACG GAA GGT TGT GAT GAA GAG GAA GTC TTA AAG TCC	432
	Phe Ser Leu Ser Thr Glu Gly Cys Asp Glu Glu Val Leu Lys Ser	
	130 135 140	
35	CTA AAG TTC TCC CTT TTC ATT GTA TGC ACT GTC ACT CTG ACT CTG TTC	480
	Leu Lys Phe Ser Leu Phe Ile Val Cys Thr Val Thr Leu Thr Leu Phe	
	145 150 155 160	
40	CTC ATG ACC ATC CTC ACA GTC ACA AAG TTC CGG GGC TTC TGT TTT ATC	528
	Leu Met Thr Ile Leu Thr Val Thr Lys Phe Arg Gly Phe Cys Phe Ile	
	165 170 175	
45	TGT TAT AAG ACA GCC CAG AGA CTG GTG TTC AAG GAC CAT CCC CAG GGC	576
	Cys Tyr Lys Thr Ala Gln Arg Leu Val Phe Lys Asp His Pro Gln Gly	
	180 185 190	
50	ACA GAA CCT GAT ATG TAC AAA TAT GAT GCC TAT TTG TGC TTC AGC AGC	624
	Thr Glu Pro Asp Met Tyr Lys Tyr Asp Ala Tyr Leu Cys Phe Ser Ser	
	195 200 205	
55	AAA GAC TTC ACA TGG GTG CAG AAT GCT TTG CTC AAA CAC CTG GAC ACT	672
	Lys Asp Phe Thr Trp Val Gln Asn Ala Leu Leu Lys His Leu Asp Thr	
	210 215 220	
60	CAA TAC AGT GAC CAA AAC AGA TTC AAC CTG TGC TTT GAA GAA AGA GAC	720
	Gln Tyr Ser Asp Gln Asn Arg Phe Asn Leu Cys Phe Glu Glu Arg Asp	
	225 230 235 240	
65	TTT GTC CCA GGA GAA AAC CGC ATT GCC AAT ATC CAG GAT GCC ATC TGG	768
	Phe Val Pro Gly Glu Asn Arg Ile Ala Asn Ile Gln Asp Ala Ile Trp	
	245 250 255	
70	AAC AGT AGA AAG ATC GTT TGT CTT GTG AGC AGA CAC TTC CTT AGA GAT	816
	Asn Ser Arg Lys Ile Val Cys Leu Val Ser Arg His Phe Leu Arg Asp	
	260 265 270	
75	GGC TGG TGC CTT GAA GCC TTC AGT TAT GCC CAG GGC AGG TGC TTA TCT	864

	Gly Trp Cys Leu Glu Ala Phe Ser Tyr Ala Gln Gly Arg Cys Leu Ser			
	275	280	285	
5	GAC CTT AAC AGT GCT CTC ATC ATG GTG GTG GTT GGG TCC TTG TCC CAG		912	
	Asp Leu Asn Ser Ala Leu Ile Met Val Val Val Gly Ser Leu Ser Gln			
	290	295	300	
10	TAC CAG TTG ATG AAA CAT CAA TCC ATC AGA GGC TTT GTA CAG AAA CAG		960	
	Tyr Gln Leu Met Lys His Gln Ser Ile Arg Gly Phe Val Gln Lys Gln			
	305	310	315	320
15	CAG TAT TTG AGG TGG CCT GAG GAT CTC CAG GAT GTT GGC TGG TTT CTT		1008	
	Gln Tyr Leu Arg Trp Pro Glu Asp Leu Gln Asp Val Gly Trp Phe Leu			
	325	330	335	
20	CAT AAA CTC TCT CAA CAG ATA CTA AAG AAA GAA AAG GAA AAG AAG AAA		1056	
	His Lys Leu Ser Gln Gln Ile Leu Lys Lys Glu Lys Lys Lys Lys Lys			
	340	345	350	
25	GAC AAT AAC ATT CCG TTG CAA ACT GTA GCA ACC ATC TCC TAATCAAAGG		1105	
	Asp Asn Asn Ile Pro Leu Gln Thr Val Ala Thr Ile Ser			
	355	360	365	
30	AGCAATTCC AACTTATCTC AAGCCACAAA TAACTCTTCA CTTTGTATTT GCACCAAGTT		1165	
	ATCATTGG GGTCCCTCTCT GGAGGTTTT TTTTCTTT TGCTACTATG AAAACAACAT		1225	
	AAATCTCTCA ATTTCGTAT CAAAAAAA AAAAAAAA TGGCGGCCGC		1275	
35	(2) INFORMATION FOR SEQ ID NO:10:			
	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 365 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
40	(ii) MOLECULE TYPE: protein			
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:			
	Cys Trp Asp Val Phe Glu Gly Leu Ser His Leu Gln Val Leu Tyr Leu			
	1 5 10 15			
50	Asn His Asn Tyr Leu Asn Ser Leu Pro Pro Gly Val Phe Ser His Leu			
	20 25 30			
	Thr Ala Leu Arg Gly Leu Ser Leu Asn Ser Asn Arg Leu Thr Val Leu			
	35 40 45			
55	Ser His Asn Asp Leu Pro Ala Asn Leu Glu Ile Leu Asp Ile Ser Arg			
	50 55 60			
60	Asn Gln Leu Leu Ala Pro Asn Pro Asp Val Phe Val Ser Leu Ser Val			
	65 70 75 80			
	Leu Asp Ile Thr His Asn Lys Phe Ile Cys Glu Cys Glu Leu Ser Thr			
	85 90 95			
65	Phe Ile Asn Trp Leu Asn His Thr Asn Val Thr Ile Ala Gly Pro Pro			
	100 105 110			

Ala Asp Ile Tyr Cys Val Tyr Pro Asp Ser Phe Ser Gly Val Ser Leu  
 115 120 125

5 Phe Ser Leu Ser Thr Glu Gly Cys Asp Glu Glu Glu Val Leu Lys Ser  
 130 135 140

10 Leu Lys Phe Ser Leu Phe Ile Val Cys Thr Val Thr Leu Thr Leu Phe  
 145 150 155 160

15 Leu Met Thr Ile Leu Thr Val Thr Lys Phe Arg Gly Phe Cys Phe Ile  
 165 170 175

20 Cys Tyr Lys Thr Ala Gln Arg Leu Val Phe Lys Asp His Pro Gln Gly  
 180 185 190

25 Thr Glu Pro Asp Met Tyr Lys Tyr Asp Ala Tyr Leu Cys Phe Ser Ser  
 195 200 205

30 Lys Asp Phe Thr Trp Val Gln Asn Ala Leu Leu Lys His Leu Asp Thr  
 210 215 220

35 Gln Tyr Ser Asp Gln Asn Arg Phe Asn Leu Cys Phe Glu Glu Arg Asp  
 225 230 235 240

40 Phe Val Pro Gly Glu Asn Arg Ile Ala Asn Ile Gln Asp Ala Ile Trp  
 245 250 255

45 Asn Ser Arg Lys Ile Val Cys Leu Val Ser Arg His Phe Leu Arg Asp  
 260 265 270

50 Gly Trp Cys Leu Glu Ala Phe Ser Tyr Ala Gln Gly Arg Cys Leu Ser  
 275 280 285

55 Asp Leu Asn Ser Ala Leu Ile Met Val Val Val Gly Ser Leu Ser Gln  
 290 295 300

Tyr Gln Leu Met Lys His Gln Ser Ile Arg Gly Phe Val Gln Lys Gln  
 305 310 315 320

60 Gln Tyr Leu Arg Trp Pro Glu Asp Leu Gln Asp Val Gly Trp Phe Leu  
 325 330 335

His Lys Leu Ser Gln Gln Ile Leu Lys Lys Glu Lys Glu Lys Lys Lys  
 340 345 350

Asp Asn Asn Ile Pro Leu Gln Thr Val Ala Thr Ile Ser  
 355 360 365

## 50 (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3138 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 60 (ix) FEATURE:

(A) NAME/KEY: CDS  
 (B) LOCATION: 1..3135

## (ix) FEATURE:

5 (A) NAME/KEY: mat\_peptide  
 (B) LOCATION: 67..3135

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

10	ATG TGG ACA CTG AAG AGA CTA ATT CTT ATC CTT TTT AAC ATA ATC CTA Met Trp Thr Leu Lys Arg Leu Ile Leu Ile Leu Phe Asn Ile Ile Leu -22 -20 -15 -10	48
15	ATT TCC AAA CTC CTT GGG GCT AGA TGG TTT CCT AAA ACT CTG CCC TGT Ile Ser Lys Leu Leu Gly Ala Arg Trp Phe Pro Lys Thr Leu Pro Cys -5 1 5 10	96
20	GAT GTC ACT CTG GAT GTT CCA AAG AAC CAT GTG ATC GTG GAC TGC ACA Asp Val Thr Leu Asp Val Pro Lys Asn His Val Ile Val Asp Cys Thr 15 20 25	144
25	GAC AAG CAT TTG ACA GAA ATT CCT GGA GGT ATT CCC ACG AAC ACC ACG Asp Lys His Leu Thr Glu Ile Pro Gly Gly Ile Pro Thr Asn Thr Thr 30 35 40	192
30	AAC CTC ACC CTC ACC ATT AAC CAC ATA CCA GAC ATC TCC CCA GCG TCC Asn Leu Thr Leu Thr Ile Asn His Ile Pro Asp Ile Ser Pro Ala Ser 45 50 55	240
35	TTT CAC AGA CTG GAC CAT CTG GTA GAG ATC GAT TTC AGA TGC AAC TGT Phe His Arg Leu Asp His Leu Val Glu Ile Asp Phe Arg Cys Asn Cys 60 65 70	288
40	GTA CCT ATT CCA CTG GGG TCA AAA AAC AAC ATG TGC ATC AAG AGG CTG Val Pro Ile Pro Leu Gly Ser Lys Asn Asn Met Cys Ile Lys Arg Leu 75 80 85 90	336
45	CAG ATT AAA CCC AGA AGC TTT AGT GGA CTC ACT TAT TTA AAA TCC CTT Gln Ile Lys Pro Arg Ser Phe Ser Gly Leu Thr Tyr Leu Lys Ser Leu 95 100 105	384
50	TAC CTG GAT GGA AAC CAG CTA CTA GAG ATA CCG CAG GGC CTC CCG CCT Tyr Leu Asp Gly Asn Gln Leu Leu Glu Ile Pro Gln Gly Leu Pro Pro 110 115 120	432
55	AGC TTA CAG CTT CTC AGC CTT GAG GCC AAC AAC ATC TTT TCC ATC AGA Ser Leu Gln Leu Leu Ser Leu Glu Ala Asn Asn Ile Phe Ser Ile Arg 125 130 135	480
60	AAA GAG AAT CTA ACA GAA CTG GCC AAC ATA GAA ATA CTC TAC CTG GGC Lys Glu Asn Leu Thr Glu Leu Ala Asn Ile Glu Ile Leu Tyr Leu Gly 140 145 150	528
65	CAA AAC TGT TAT TAT CGA AAT CCT TGT TAT GTT TCA TAT TCA ATA GAG Gln Asn Cys Tyr Tyr Arg Asn Pro Cys Tyr Val Ser Tyr Ser Ile Glu 155 160 165 170	576
70	AAA GAT GCC TTC CTA AAC TTG ACA AAG TTA AAA GTG CTC TCC CTG AAA Lys Asp Ala Phe Leu Asn Leu Thr Lys Leu Lys Val Leu Ser Leu Lys 175 180 185	624

	GAT AAC AAT GTC ACA GCC GTC CCT ACT GTT TTG CCA TCT ACT TTA ACA Asp Asn Asn Val Thr Ala Val Pro Thr Val Leu Pro Ser Thr Leu Thr 190 195 200	672
5	GAA CTA TAT CTC TAC AAC AAC ATG ATT GCA AAA ATC CAA GAA GAT GAT Glu Leu Tyr Leu Tyr Asn Asn Met Ile Ala Lys Ile Gln Glu Asp Asp 205 210 215	720
10	TTT AAT AAC CTC AAC CAA TTA CAA ATT CTT GAC CTA AGT GGA AAT TGC. Phe Asn Asn Leu Asn Gln Leu Gln Ile Leu Asp Leu Ser Gly Asn Cys 220 225 230	768
15	CCT CGT TGT TAT AAT GCC CCA TTT CCT TGT GCG CCG TGT AAA AAT AAT Pro Arg Cys Tyr Asn Ala Pro Phe Pro Cys Ala Pro Cys Lys Asn Asn 235 240 245 250	816
20	TCT CCC CTA CAG ATC CCT GTA AAT GCT TTT GAT GCG CTG ACA GAA TTA Ser Pro Leu Gln Ile Pro Val Asn Ala Phe Asp Ala Leu Thr Glu Leu 255 260 265	864
25	AAA GTT TTA CGT CTA CAC AGT AAC TCT CTT CAG CAT GTG CCC CCA AGA Lys Val Leu Arg Leu His Ser Asn Ser Leu Gln His Val Pro Pro Arg 270 275 280	912
30	TGG TTT AAG AAC ATC AAC AAA CTC CAG GAA CTG GAT CTG TCC CAA AAC Trp Phe Lys Asn Ile Asn Lys Leu Gln Glu Leu Asp Leu Ser Gln Asn 285 290 295	960
35	TTC TTG GCC AAA GAA ATT GGG GAT GCT AAA TTT CTG CAT TTT CTC CCC Phe Leu Ala Lys Glu Ile Gly Asp Ala Lys Phe Leu His Phe Leu Pro 300 305 310	1008
40	AGC CTC ATC CAA TTG GAT CTG TCT TTC AAT TTT GAA CTT CAG GTC TAT Ser Leu Ile Gln Leu Asp Leu Ser Phe Asn Phe Glu Leu Gln Val Tyr 315 320 325 330	1056
45	CGT GCA TCT ATG AAT CTA TCA CAA GCA TTT TCT TCA CTG AAA AGC CTG Arg Ala Ser Met Asn Leu Ser Gln Ala Phe Ser Ser Leu Lys Ser Leu 335 340 345	1104
50	AAA ATT CTG CGG ATC AGA GGA TAT GTC TTT AAA GAG TTG AAA AGC TTT Lys Ile Leu Arg Ile Arg Gly Tyr Val Phe Lys Glu Leu Lys Ser Phe 350 355 360	1152
55	AAC CTC TCG CCA TTA CAT AAT CTT CAA AAT CTT GAA GTT CTT GAT CTT Asn Leu Ser Pro Leu His Asn Leu Gln Asn Leu Glu Val Leu Asp Leu 365 370 375	1200
60	GGC ACT AAC TTT ATA AAA ATT GCT AAC CTC AGC ATG TTT AAA CAA TTT Gly Thr Asn Phe Ile Lys Ile Ala Asn Leu Ser Met Phe Lys Gln Phe 380 385 390	1248
	AAA AGA CTG AAA GTC ATA GAT CTT TCA GTG AAT AAA ATA TCA CCT TCA Lys Arg Leu Lys Val Ile Asp Leu Ser Val Asn Lys Ile Ser Pro Ser 395 400 405 410	1296
	GGA GAT TCA AGT GAA GTT GGC TTC TGC TCA AAT GCC AGA ACT TCT GTA Gly Asp Ser Ser Glu Val Gly Phe Cys Ser Asn Ala Arg Thr Ser Val 415 420 425	1344

	GAA AGT TAT GAA CCC CAG GTC CTG GAA CAA TTA CAT TAT TTC AGA TAT Glu Ser Tyr Glu Pro Gln Val Leu Glu Gln Leu His Tyr Phe Arg Tyr 430 435 440	1392
5	GAT AAG TAT GCA AGG AGT TGC AGA TTC AAA AAC AAA GAG GCT TCT TTC Asp Lys Tyr Ala Arg Ser Cys Arg Phe Lys Asn Lys Glu Ala Ser Phe 445 450 455	1440
10	ATG TCT GTT AAT GAA AGC TGC TAC AAG TAT GGG CAG ACC TTG GAT CTA Met Ser Val Asn Glu Ser Cys Tyr Lys Tyr Gly Gln Thr Leu Asp Leu 460 465 470	1488
15	AGT AAA AAT AGT ATA TTT TTT GTC AAG TCC TCT GAT TTT CAG CAT CTT Ser Lys Asn Ser Ile Phe Phe Val Lys Ser Ser Asp Phe Gln His Leu 475 480 485 490	1536
20	TCT TTC CTC AAA TGC CTG AAT CTG TCA GGA AAT CTC ATT AGC CAA ACT Ser Phe Leu Lys Cys Leu Asn Leu Ser Gly Asn Leu Ile Ser Gln Thr 495 500 505	1584
25	CTT AAT GGC AGT GAA TTC CAA CCT TTA GCA GAG CTG AGA TAT TTG GAC Leu Asn Gly Ser Glu Phe Gln Pro Leu Ala Glu Leu Arg Tyr Leu Asp 510 515 520	1632
30	TTC TCC AAC AAC CGG CTT GAT TTA CTC CAT TCA ACA GCA TTT GAA GAG Phe Ser Asn Asn Arg Leu Asp Leu Leu His Ser Thr Ala Phe Glu Glu 525 530 535	1680
35	CTT CAC AAA CTG GAA GTT CTG GAT ATA AGC AGT AAT AGC CAT TAT TTT Leu His Lys Leu Glu Val Leu Asp Ile Ser Ser Asn Ser His Tyr Phe 540 545 550	1728
40	CAA TCA GAA GGA ATT ACT CAT ATG CTA AAC TTT ACC AAG AAC CTA AAG Gln Ser Glu Gly Ile Thr His Met Leu Asn Phe Thr Lys Asn Leu Lys 555 560 565 570	1776
45	GTT CTG CAG AAA CTG ATG ATG AAC GAC AAT GAC ATC TCT TCC TCC ACC Val Leu Gln Lys Leu Met Met Asn Asp Asn Asp Ile Ser Ser Ser Thr 575 580 585	1824
50	AGC AGG ACC ATG GAG AGT GAG TCT CTT AGA ACT CTG GAA TTC AGA GGA Ser Arg Thr Met Glu Ser Glu Ser Leu Arg Thr Leu Glu Phe Arg Gly 590 595 600	1872
55	AAT CAC TTA GAT GTT TTA TGG AGA GAA GGT GAT AAC AGA TAC TTA CAA Asn His Leu Asp Val Leu Trp Arg Glu Gly Asp Asn Arg Tyr Leu Gln 605 610 615	1920
60	TTA TTC AAG AAT CTG CTA AAA TTA GAG GAA TTA GAC ATC TCT AAA AAT Leu Phe Lys Asn Leu Leu Lys Leu Glu Glu Leu Asp Ile Ser Lys Asn 620 625 630	1968
	TCC CTA AGT TTC TTG CCT TCT GGA GTT TTT GAT GGT ATG CCT CCA AAT Ser Leu Ser Phe Leu Pro Ser Gly Val Phe Asp Gly Met Pro Pro Asn 635 640 645 650	2016
	CTA AAG AAT CTC TCT TTG GCC AAA AAT GGG CTC AAA TCT TTC AGT TGG Leu Lys Asn Leu Ser Leu Ala Lys Asn Gly Leu Lys Ser Phe Ser Trp 655 660 665	2064
	AAG AAA CTC CAG TGT CTA AAG AAC CTG GAA ACT TTG GAC CTC AGC CAC	2112

	Lys Lys Leu Gln Cys Leu Lys Asn Leu Glu Thr Leu Asp Leu Ser His 670 675 680	
5	AAC CAA CTG ACC ACT GTC CCT GAG AGA TTA TCC AAC TGT TCC AGA AGC Asn Gln Leu Thr Thr Val Pro Glu Arg Leu Ser Asn Cys Ser Arg Ser 685 690 695	2160
10	CTC AAG AAT CTG ATT CTT AAG AAT AAT CAA ATC AGG AGT CTG ACG AAG Leu Lys Asn Leu Ile Leu Lys Asn Asn Gln Ile Arg Ser Leu Thr Lys 700 705 710	2208
15	TAT TTT CTA CAA GAT GCC TTC CAG TTG CGA TAT CTG GAT CTC AGC TCA Tyr Phe Leu Gln Asp Ala Phe Gln Leu Arg Tyr Leu Asp Leu Ser Ser 715 720 725 730	2256
20	AAT AAA ATC CAG ATG ATC CAA AAG ACC AGC TTC CCA GAA AAT GTC CTC Asn Lys Ile Gln Met Ile Gln Lys Thr Ser Phe Pro Glu Asn Val Leu 735 740 745	2304
25	AAC AAT CTG AAG ATG TTG CTT TTG CAT CAT AAT CGG TTT CTG TGC ACC Asn Asn Leu Lys Met Leu Leu His His Asn Arg Phe Leu Cys Thr 750 755 760	2352
30	TGT GAT GCT GTG TGG TTT GTC TGG TGG GTT AAC CAT ACG GAG GTG ACT Cys Asp Ala Val Trp Phe Val Trp Trp Val Asn His Thr Glu Val Thr 765 770 775	2400
35	ATT CCT TAC CTG GCC ACA GAT GTG ACT TGT GTG GGG CCA GGA GCA CAC Ile Pro Tyr Leu Ala Thr Asp Val Thr Cys Val Gly Pro Gly Ala His 780 785 790	2448
40	AAG GGC CAA AGT GTG ATC TCC CTG GAT CTG TAC ACC TGT GAG TTA GAT Lys Gly Gln Ser Val Ile Ser Leu Asp Leu Tyr Thr Cys Glu Leu Asp 795 800 805 810	2496
45	CTG ACT AAC CTG ATT CTG TTC TCA CTT TCC ATA TCT GTA TCT CTC TTT Leu Thr Asn Leu Ile Leu Phe Ser Leu Ser Ile Ser Val Ser Leu Phe 815 820 825	2544
50	CTC ATG GTG ATG ATG ACA GCA AGT CAC CTC TAT TTC TGG GAT GTG TGG Leu Met Val Met Met Thr Ala Ser His Leu Tyr Phe Trp Asp Val Trp 830 835 840	2592
55	TAT ATT TAC CAT TTC TGT AAG GCC AAG ATA AAG GGG TAT CAG CGT CTA Tyr Ile Tyr His Phe Cys Lys Ala Lys Ile Lys Gly Tyr Gln Arg Leu 845 850 855	2640
60	ATA TCA CCA GAC TGT TGC TAT GAT GCT TTT ATT GTG TAT GAC ACT AAA Ile Ser Pro Asp Cys Cys Tyr Asp Ala Phe Ile Val Tyr Asp Thr Lys 860 865 870	2688
	GAC CCA GCT GTG ACC GAG TGG GTT TTG GCT GAG CTG GTG GCC AAA CTG Asp Pro Ala Val Thr Glu Trp Val Leu Ala Glu Leu Val Ala Lys Leu 875 880 885 890	2736
	GAA GAC CCA AGA GAG AAA CAT TTT AAT TTA TGT CTC GAG GAA AGG GAC Glu Asp Pro Arg Glu Lys His Phe Asn Leu Cys Leu Glu Arg Asp 895 900 905	2784
	TGG TTA CCA GGG CAG CCA GTT CTG GAA AAC CTT TCC CAG AGC ATA CAG Trp Leu Pro Gly Gln Pro Val Leu Glu Asn Leu Ser Gln Ser Ile Gln	2832

	910	915	920	
5	CTT AGC AAA AAG ACA GTG TTT GTG ATG ACA GAC AAG TAT GCA AAG ACT Leu Ser Lys Lys Thr Val Phe Val Met Thr Asp Lys Tyr Ala Lys Thr 925	930	935	2880
10	GAA AAT TTT AAG ATA GCA TTT TAC TTG TCC CAT CAG AGG CTC ATG GAT Glu Asn Phe Lys Ile Ala Phe Tyr Leu Ser His Gln Arg Leu Met Asp 940	945	950	2928
15	GAA AAA GTT GAT GTG ATT ATC TTG ATA TTT CTT GAG AAG CCC TTT CAG Glu Lys Val Asp Val Ile Ile Leu Ile Phe Leu Glu Lys Pro Phe Gln 955	960	965	2976
20	AAG TCC AAG TTC CTC CAG CTC CGG AAA AGG CTC TGT GGG AGT TCT GTC Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg Leu Cys Gly Ser Ser Val 975	980	985	3024
25	CTT GAG TGG CCA ACA AAC CCG CAA GCT CAC CCA TAC TTC TGG CAG TGT Leu Glu Trp Pro Thr Asn Pro Gln Ala His Pro Tyr Phe Trp Gln Cys 990	995	1000	3072
30	CTA AAG AAC GCC CTG GCC ACA GAC AAT CAT GTG GCC TAT AGT CAG GTG Leu Lys Asn Ala Leu Ala Thr Asp Asn His Val Ala Tyr Ser Gln Val 1005	1010	1015	3120
35	TTC AAG GAA ACG GTC TAG Phe Lys Glu Thr Val 1020			3138
40	(2) INFORMATION FOR SEQ ID NO:12:			
45	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1045 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear			
50	(ii) MOLECULE TYPE: protein			
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:			
60	Met Trp Thr Leu Lys Arg Leu Ile Leu Ile Leu Phe Asn Ile Ile Leu -22 -20 -15 -10			
65	Ile Ser Lys Leu Leu Gly Ala Arg Trp Phe Pro Lys Thr Leu Pro Cys -5 1 5 10			
70	Asp Val Thr Leu Asp Val Pro Lys Asn His Val Ile Val Asp Cys Thr 15 20 25			
75	Asp Lys His Leu Thr Glu Ile Pro Gly Gly Ile Pro Thr Asn Thr Thr 30 35 40			
80	Asn Leu Thr Leu Thr Ile Asn His Ile Pro Asp Ile Ser Pro Ala Ser 45 50 55			
85	Phe His Arg Leu Asp His Leu Val Glu Ile Asp Phe Arg Cys Asn Cys 60 65 70			
90	Val Pro Ile Pro Leu Gly Ser Lys Asn Asn Met Cys Ile Lys Arg Leu			

	75	80	85	90
	Gln Ile Lys Pro Arg Ser Phe Ser Gly Leu Thr Tyr Leu Lys Ser Leu			
	95		100	105
5	Tyr Leu Asp Gly Asn Gln Leu Leu Glu Ile Pro Gln Gly Leu Pro Pro			
	110		115	120
	Ser Leu Gln Leu Leu Ser Leu Glu Ala Asn Asn Ile Phe Ser Ile Arg			
10	125		130	135
	Lys Glu Asn Leu Thr Glu Leu Ala Asn Ile Glu Ile Leu Tyr Leu Gly			
	140		145	150
15	Gln Asn Cys Tyr Tyr Arg Asn Pro Cys Tyr Val Ser Tyr Ser Ile Glu			
	155		160	165
	Lys Asp Ala Phe Leu Asn Leu Thr Lys Leu Lys Val Leu Ser Leu Lys			
20	175		180	185
	Asp Asn Asn Val Thr Ala Val Pro Thr Val Leu Pro Ser Thr Leu Thr			
	190		195	200
	Glu Leu Tyr Leu Tyr Asn Asn Met Ile Ala Lys Ile Gln Glu Asp Asp			
25	205		210	215
	Phe Asn Asn Leu Asn Gln Leu Gln Ile Leu Asp Leu Ser Gly Asn Cys			
	220		225	230
30	Pro Arg Cys Tyr Asn Ala Pro Phe Pro Cys Ala Pro Cys Lys Asn Asn			
	235		240	245
	Ser Pro Leu Gln Ile Pro Val Asn Ala Phe Asp Ala Leu Thr Glu Leu			
	255		260	265
35	Lys Val Leu Arg Leu His Ser Asn Ser Leu Gln His Val Pro Pro Arg			
	270		275	280
	Trp Phe Lys Asn Ile Asn Lys Leu Gln Glu Leu Asp Leu Ser Gln Asn			
40	285		290	295
	Phe Leu Ala Lys Glu Ile Gly Asp Ala Lys Phe Leu His Phe Leu Pro			
	300		305	310
45	Ser Leu Ile Gln Leu Asp Leu Ser Phe Asn Phe Glu Leu Gln Val Tyr			
	315		320	325
	Arg Ala Ser Met Asn Leu Ser Gln Ala Phe Ser Ser Leu Lys Ser Leu			
50	335		340	345
	Lys Ile Leu Arg Ile Arg Gly Tyr Val Phe Lys Glu Leu Lys Ser Phe			
	350		355	360
	Asn Leu Ser Pro Leu His Asn Leu Gln Asn Leu Glu Val Leu Asp Leu			
55	365		370	375
	Gly Thr Asn Phe Ile Lys Ile Ala Asn Leu Ser Met Phe Lys Gln Phe			
	380		385	390
60	Lys Arg Leu Lys Val Ile Asp Leu Ser Val Asn Lys Ile Ser Pro Ser			
	395		400	405
				410

	Gly Asp Ser Ser Glu Val Gly Phe Cys Ser Asn Ala Arg Thr Ser Val			
	415	420	425	
5	Glu Ser Tyr Glu Pro Gln Val Leu Glu Gln Leu His Tyr Phe Arg Tyr			
	430	435	440	
	Asp Lys Tyr Ala Arg Ser Cys Arg Phe Lys Asn Lys Glu Ala Ser Phe			
	445	450	455	
10	Met Ser Val Asn Glu Ser Cys Tyr Lys Tyr Gly Gln Thr Leu Asp Leu			
	460	465	470	
15	Ser Lys Asn Ser Ile Phe Phe Val Lys Ser Ser Asp Phe Gln His Leu			
	475	480	485	490
	Ser Phe Leu Lys Cys Leu Asn Leu Ser Gly Asn Leu Ile Ser Gln Thr			
	495	500	505	
20	Leu Asn Gly Ser Glu Phe Gln Pro Leu Ala Glu Leu Arg Tyr Leu Asp			
	510	515	520	
	Phe Ser Asn Asn Arg Leu Asp Leu Leu His Ser Thr Ala Phe Glu Glu			
25	525	530	535	
	Leu His Lys Leu Glu Val Leu Asp Ile Ser Ser Asn Ser His Tyr Phe			
	540	545	550	
30	Gln Ser Glu Gly Ile Thr His Met Leu Asn Phe Thr Lys Asn Leu Lys			
	555	560	565	570
	Val Leu Gln Lys Leu Met Met Asn Asp Asn Asp Ile Ser Ser Ser Thr			
	575	580	585	
35	Ser Arg Thr Met Glu Ser Glu Ser Leu Arg Thr Leu Glu Phe Arg Gly			
	590	595	600	
	Asn His Leu Asp Val Leu Trp Arg Glu Gly Asp Asn Arg Tyr Leu Gln			
40	605	610	615	
	Leu Phe Lys Asn Leu Leu Lys Leu Glu Glu Leu Asp Ile Ser Lys Asn			
	620	625	630	
45	Ser Leu Ser Phe Leu Pro Ser Gly Val Phe Asp Gly Met Pro Pro Asn			
	635	640	645	650
	Leu Lys Asn Leu Ser Leu Ala Lys Asn Gly Leu Lys Ser Phe Ser Trp			
	655	660	665	
50	Lys Lys Leu Gln Cys Leu Lys Asn Leu Glu Thr Leu Asp Leu Ser His			
	670	675	680	
	Asn Gln Leu Thr Thr Val Pro Glu Arg Leu Ser Asn Cys Ser Arg Ser			
55	685	690	695	
	Leu Lys Asn Leu Ile Leu Lys Asn Asn Gln Ile Arg Ser Leu Thr Lys			
	700	705	710	
60	Tyr Phe Leu Gln Asp Ala Phe Gln Leu Arg Tyr Leu Asp Leu Ser Ser			
	715	720	725	730

	Asn Lys Ile Gln Met Ile Gln Lys Thr Ser Phe Pro Glu Asn Val Leu			
	735	740	745	
5	Asn Asn Leu Lys Met Leu Leu Leu His His Asn Arg Phe Leu Cys Thr			
	750	755	760	
	Cys Asp Ala Val Trp Phe Val Trp Trp Val Asn His Thr Glu Val Thr			
	765	770	775	
10	Ile Pro Tyr Leu Ala Thr Asp Val Thr Cys Val Gly Pro Gly Ala His.			
	780	785	790	
	Lys Gly Gln Ser Val Ile Ser Leu Asp Leu Tyr Thr Cys Glu Leu Asp			
	795	800	805	810
15	Leu Thr Asn Leu Ile Leu Phe Ser Leu Ser Ile Ser Val Ser Leu Phe			
	815	820	825	
20	Leu Met Val Met Met Thr Ala Ser His Leu Tyr Phe Trp Asp Val Trp			
	830	835	840	
	Tyr Ile Tyr His Phe Cys Lys Ala Lys Ile Lys Gly Tyr Gln Arg Leu			
	845	850	855	
25	Ile Ser Pro Asp Cys Cys Tyr Asp Ala Phe Ile Val Tyr Asp Thr Lys			
	860	865	870	
	Asp Pro Ala Val Thr Glu Trp Val Leu Ala Glu Leu Val Ala Lys Leu			
	875	880	885	890
30	Glu Asp Pro Arg Glu Lys His Phe Asn Leu Cys Leu Glu Glu Arg Asp			
	895	900	905	
35	Trp Leu Pro Gly Gln Pro Val Leu Glu Asn Leu Ser Gln Ser Ile Gln			
	910	915	920	
	Leu Ser Lys Lys Thr Val Phe Val Met Thr Asp Lys Tyr Ala Lys Thr			
	925	930	935	
40	Glu Asn Phe Lys Ile Ala Phe Tyr Leu Ser His Gln Arg Leu Met Asp			
	940	945	950	
	Glu Lys Val Asp Val Ile Ile Leu Ile Phe Leu Glu Lys Pro Phe Gln			
	955	960	965	970
45	Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg Leu Cys Gly Ser Ser Val			
	975	980	985	
50	Leu Glu Trp Pro Thr Asn Pro Gln Ala His Pro Tyr Phe Trp Gln Cys			
	990	995	1000	
	Leu Lys Asn Ala Leu Ala Thr Asp Asn His Val Ala Tyr Ser Gln Val			
	1005	1010	1015	
55	Phe Lys Glu Thr Val			
	1020			

## (2) INFORMATION FOR SEQ ID NO:13:

- 60           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 180 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5           (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..177

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

15	CTT GGA AAA CCT CTT CAG AAG TCT AAG TTT CTT CAG CTC AGG AAG AGA Leu Gly Lys Pro Leu Gln Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg 1                 5                             10                     15	48
20	CTC TGC AGG AGC TCT GTC CTT GAG TGG CCT GCA AAT CCA CAG GCT CAC Leu Cys Arg Ser Ser Val Leu Glu Trp Pro Ala Asn Pro Gln Ala His 20                 25                             30	96
25	CCA TAC TTC TGG CAG TGC CTG AAA AAT GCC CTG ACC ACA GAC AAT CAT Pro Tyr Phe Trp Gln Cys Leu Lys Asn Ala Leu Thr Thr Asp Asn His 35                 40                             45	144
30	GTG GCT TAT AGT CAA ATG TTC AAG GAA ACA GTC TAG Val Ala Tyr Ser Gln Met Phe Lys Glu Thr Val 50                 55	180

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40           (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

45	Leu Gly Lys Pro Leu Gln Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg 1                 5                             10                     15
50	Leu Cys Arg Ser Ser Val Leu Glu Trp Pro Ala Asn Pro Gln Ala His 20                 25                             30
55	Pro Tyr Phe Trp Gln Cys Leu Lys Asn Ala Leu Thr Thr Asp Asn His 35                 40                             45
60	Val Ala Tyr Ser Gln Met Phe Lys Glu Thr Val 50                 55

55           (2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## 5 (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..988

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

	G AAT TCC AGA CTT ATA AAC TTG AAA AAT CTC TAT TTG GCC TGG AAC	46
	Asn Ser Arg Leu Ile Asn Leu Lys Asn Leu Tyr Leu Ala Trp Asn	
15	1 5 10 15	
	TGC TAT TTT AAC AAA GTT TGC GAG AAA ACT AAC ATA GAA GAT GGA GTA	94
	Cys Tyr Phe Asn Lys Val Cys Glu Lys Thr Asn Ile Glu Asp Gly Val	
	20 25 30	
20	TTT GAA ACG CTG ACA AAT TTG GAG TTG CTA TCA CTA TCT TTC AAT TCT	142
	Phe Glu Thr Leu Thr Asn Leu Glu Leu Leu Ser Leu Ser Phe Asn Ser	
	35 40 45	
25	CTT TCA CAT GTG CCA CCC AAA CTG CCA AGC TCC CTA CGC AAA CTT TTT	190
	Leu Ser His Val Pro Pro Lys Leu Pro Ser Ser Leu Arg Lys Leu Phe	
	50 55 60	
30	CTG AGC AAC ACC CAG ATC AAA TAC ATT AGT GAA GAA GAT TTC AAG GGA	238
	Leu Ser Asn Thr Gln Ile Lys Tyr Ile Ser Glu Glu Asp Phe Lys Gly	
	65 70 75	
35	TTG ATA AAT TTA ACA TTA CTA GAT TTA AGC GGG AAC TGT CCG AGG TGC	286
	Leu Ile Asn Leu Thr Leu Leu Asp Leu Ser Gly Asn Cys Pro Arg Cys	
	80 85 90 95	
	TTC AAT GCC CCA TTT CCA TGC GTG CCT TGT GAT GGT GGT GCT TCA ATT	334
	Phe Asn Ala Pro Phe Pro Cys Val Pro Cys Asp Gly Gly Ala Ser Ile	
	100 105 110	
40	AAT ATA GAT CGT TTT GCT TTT CAA AAC TTG ACC CAA CTT CGA TAC CTA	382
	Asn Ile Asp Arg Phe Ala Phe Gln Asn Leu Thr Gln Leu Arg Tyr Leu	
	115 120 125	
45	AAC CTC TCT AGC ACT TCC CTC AGG AAG ATT AAT GCT GCC TGG TTT AAA	430
	Asn Leu Ser Ser Thr Ser Leu Arg Lys Ile Asn Ala Ala Trp Phe Lys	
	130 135 140	
50	AAT ATG CCT CAT CTG AAG GTG CTG GAT CTT GAA TTC AAC TAT TTA GTG	478
	Asn Met Pro His Leu Lys Val Leu Asp Leu Glu Phe Asn Tyr Leu Val	
	145 150 155	
55	GGA GAA ATA GCC TCT GGG GCA TTT TTA ACG ATG CTG CCC CGC TTA GAA	526
	Gly Glu Ile Ala Ser Gly Ala Phe Leu Thr Met Leu Pro Arg Leu Glu	
	160 165 170 175	
	ATA CTT GAC TTG TCT TTT AAC TAT ATA AAG GGG AGT TAT CCA CAG CAT	574
	Ile Leu Asp Leu Ser Phe Asn Tyr Ile Lys Gly Ser Tyr Pro Gln His	
	180 185 190	
60	ATT AAT ATT TCC AGA AAC TTC TCT AAA CTT TTG TCT CTA CGG GCA TTG	622
	Ile Asn Ile Ser Arg Asn Phe Ser Lys Leu Leu Ser Leu Arg Ala Leu	

	195	200	205		
	CAT TTA AGA GGT TAT GTG TTC CAG GAA CTC AGA GAA GAT GAT TTC CAG				
5	His Leu Arg Gly Tyr Val Phe Gln Glu Leu Arg Glu Asp Asp Phe Gln	210	215	220	
	CCC CTG ATG CAG CTT CCA AAC TTA TCG ACT ATC AAC TTG GGT ATT AAT				
	Pro Leu Met Gln Leu Pro Asn Leu Ser Thr Ile Asn Leu Gly Ile Asn	225	230	235	
10	Phe Ile Lys Gln Ile Asp Phe Lys Leu Phe Gln Asn Phe Ser Asn Leu	240	245	250	
	TTC ATT AAG CAA ATC GAT TTC AAA CTT TTC CAA AAT TTC TCC AAT CTG				
	Phe Ile Lys Gln Ile Asp Phe Lys Leu Phe Gln Asn Phe Ser Asn Leu	255			
15	GAA ATT ATT TAC TTG TCA GAA AAC AGA ATA TCA CCG TTG GTA AAA GAT				
	Glu Ile Ile Tyr Leu Ser Glu Asn Arg Ile Ser Pro Leu Val Lys Asp	260	265	270	
20	ACC CGG CAG AGT TAT GCA AAT AGT TCC TCT TTT CAA CGT CAT ATC CGG				
	Thr Arg Gln Ser Tyr Ala Asn Ser Ser Phe Gln Arg His Ile Arg	275	280	285	
25	AAA CGA CGC TCA ACA GAT TTT GAG TTT GAC CCA CAT TCG AAC TTT TAT				
	Lys Arg Arg Ser Thr Asp Phe Glu Phe Asp Pro His Ser Asn Phe Tyr	290	295	300	
	CAT TTC ACC CGT CCT TTA ATA AAG CCA CAA TGT GCT GCT TAT GGA AAA				
	His Phe Thr Arg Pro Leu Ile Lys Pro Gln Cys Ala Ala Tyr Gly Lys	305	310	315	
30	GCC TTA GAT TTA AGC CTC AAC AGT ATT TTC TT				
	Ala Leu Asp Leu Ser Leu Asn Ser Ile Phe	320	325		
35	(2) INFORMATION FOR SEQ ID NO:16:				
	(i) SEQUENCE CHARACTERISTICS:				
40	(A) LENGTH: 329 amino acids				
	(B) TYPE: amino acid				
	(D) TOPOLOGY: linear				
	(ii) MOLECULE TYPE: protein				
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:				
	Asn Ser Arg Leu Ile Asn Leu Lys Asn Leu Tyr Leu Ala Trp Asn Cys				
	1	5	10	15	
50	Tyr Phe Asn Lys Val Cys Glu Lys Thr Asn Ile Glu Asp Gly Val Phe				
	20	25	30		
	Glu Thr Leu Thr Asn Leu Glu Leu Leu Ser Leu Ser Phe Asn Ser Leu				
55	35	40	45		
	Ser His Val Pro Pro Lys Leu Pro Ser Ser Leu Arg Lys Leu Phe Leu				
	50	55	60		
60	Ser Asn Thr Gln Ile Lys Tyr Ile Ser Glu Glu Asp Phe Lys Gly Leu				
	65	70	75	80	

Ile Asn Leu Thr Leu Leu Asp Leu Ser Gly Asn Cys Pro Arg Cys Phe  
 85 90 95

5 Asn Ala Pro Phe Pro Cys Val Pro Cys Asp Gly Gly Ala Ser Ile Asn  
 100 105 110

Ile Asp Arg Phe Ala Phe Gln Asn Leu Thr Gln Leu Arg Tyr Leu Asn  
 115 120 125

10 Leu Ser Ser Thr Ser Leu Arg Lys Ile Asn Ala Ala Trp Phe Lys Asn.  
 130 135 140

Met Pro His Leu Lys Val Leu Asp Leu Glu Phe Asn Tyr Leu Val Gly  
 145 150 155 160

15 Glu Ile Ala Ser Gly Ala Phe Leu Thr Met Leu Pro Arg Leu Glu Ile  
 165 170 175

20 Leu Asp Leu Ser Phe Asn Tyr Ile Lys Gly Ser Tyr Pro Gln His Ile  
 180 185 190

Asn Ile Ser Arg Asn Phe Ser Lys Leu Leu Ser Leu Arg Ala Leu His  
 195 200 205

25 Leu Arg Gly Tyr Val Phe Gln Glu Leu Arg Glu Asp Asp Phe Gln Pro  
 210 215 220

Leu Met Gln Leu Pro Asn Leu Ser Thr Ile Asn Leu Gly Ile Asn Phe  
 225 230 235 240

30 Ile Lys Gln Ile Asp Phe Lys Leu Phe Gln Asn Phe Ser Asn Leu Glu  
 245 250 255

35 Ile Ile Tyr Leu Ser Glu Asn Arg Ile Ser Pro Leu Val Lys Asp Thr  
 260 265 270

Arg Gln Ser Tyr Ala Asn Ser Ser Ser Phe Gln Arg His Ile Arg Lys  
 275 280 285

40 Arg Arg Ser Thr Asp Phe Glu Phe Asp Pro His Ser Asn Phe Tyr His  
 290 295 300

Phe Thr Arg Pro Leu Ile Lys Pro Gln Cys Ala Ala Tyr Gly Lys Ala  
 305 310 315 320

45 Leu Asp Leu Ser Leu Asn Ser Ile Phe  
 325

50 (2) INFORMATION FOR SEQ ID NO:17:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1557 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: cDNA

60 (ix) FEATURE:  
 (A) NAME/KEY: CDS

(B) LOCATION: 1..513

## (ix) FEATURE:

(A) NAME/KEY: misc\_feature

5 (B) LOCATION: 278

(D) OTHER INFORMATION: /note= "nucleotide 278 designated G, may be G or C"

## (ix) FEATURE:

10 (A) NAME/KEY: misc\_feature

(B) LOCATION: 445

(D) OTHER INFORMATION: /note= "nucleotide 445 designated A, may be A or T"

## 15 (ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION: 572

(D) OTHER INFORMATION: /note= "nucleotides 572, 593, 600, 607, 617, 622, 625, 631, 640, 646, 653, 719, 775, and 861 are 20 designated C; each may be A, C, G, or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

25	CAG TCT CTT TCC ACA TCC CAA ACT TTC TAT GAT GCT TAC ATT TCT TAT Gln Ser Leu Ser Thr Ser Gln Thr Phe Tyr Asp Ala Tyr Ile Ser Tyr 1 5 10 15	48
30	GAC ACC AAA GAT GCC TCT GTT ACT GAC TGG GTG ATA AAT GAG CTG CGC Asp Thr Lys Asp Ala Ser Val Thr Asp Trp Val Ile Asn Glu Leu Arg 20 25 30	96
35	TAC CAC CTT GAA GAG AGC CGA GAC AAA AAC GTT CTC CTT TGT CTA GAG Tyr His Leu Glu Ser Arg Asp Lys Asn Val Leu Leu Cys Leu Glu 35 40 45	144
40	GAG AGG GAT TGG GAC CCG GGA TTG GCC ATC ATC GAC AAC CTC ATG CAG Glu Arg Asp Trp Asp Pro Gly Leu Ala Ile Ile Asp Asn Leu Met Gln 50 55 60	192
45	AGC ATC AAC CAA AGC AAG AAA ACA GTA TTT GTT TTA ACC AAA AAA TAT Ser Ile Asn Gln Ser Lys Lys Thr Val Phe Val Leu Thr Lys Lys Tyr 65 70 75 80	240
50	GCA AAA AGC TGG AAC TTT AAA ACA GCT TTT TAC TTG GGC TTG CAG AGG Ala Lys Ser Trp Asn Phe Lys Thr Ala Phe Tyr Leu Gly Leu Gln Arg 85 90 95	288
55	CTA ATG GGT GAG AAC ATG GAT GTG ATT ATA TTT ATC CTG CTG GAG CCA Leu Met Gly Glu Asn Met Asp Val Ile Ile Phe Ile Leu Leu Glu Pro 100 105 110	336
60	GTG TTA CAG CAT TCT CCG TAT TTG AGG CTA CGG CAG CGG ATC TGT AAG Val Leu Gln His Ser Pro Tyr Leu Arg Leu Arg Gln Arg Ile Cys Lys 115 120 125	384
65	AGC TCC ATC CTC CAG TGG CCT GAC AAC CCG AAG GCA GAA AGG TTG TTT Ser Ser Ile Leu Gln Trp Pro Asp Asn Pro Lys Ala Glu Arg Leu Phe 130 135 140	432
70	TGG CAA ACT CTG AGA AAT GTG GTC TTG ACT GAA AAT GAT TCA CGG TAT	480

	Trp Gln Thr Leu Arg Asn Val Val Leu Thr Glu Asn Asp Ser Arg Tyr		
145	150	155	160
5	AAC AAT ATG TAT GTC GAT TCC ATT AAG CAA TAC TAACTGACGT TAAGTCATGA Asn Asn Met Tyr Val Asp Ser Ile Lys Gln Tyr		533
	165	170	
10	TTTCGCGCCA TAATAAAGAT GCAAAGGAAT GACATTTCCG TATTAGTTAT CTATTGCTAC		593
15	GGTAACCAAA TTACTCCCAA AAACCTTACG TCGGTTCAA AACAAACCACA TTCTGCTGGC CCCACAGTTT TTGAGGGTCA GGAGTCCAGG CCCAGCATAA CTGGGTCTTC TGCTTCAGGG TGTCTCCAGA GGCTGCAATG TAGGTGTTCA CCAGAGACAT AGGCATCACT GGGGTCACAC		653
20	TCCATGTGGT TGTTTCTGG ATTCAATTCC TCCTGGGCTA TTGGCCAAAG GCTATACTCA TGTAAGCCAT GCGAGCCTAT CCCACAACGG CAGCTTGCTT CATCAGAGCT AGCAAAAAAG		773
25	AGAGGTTGCT AGCAAGATGA AGTCACAATC TTTTGTAAATC GAATCAAAAA AGTGATATCT CATCACTTTG GCCATATTCT ATTTGTTAGA AGTAAACCAC AGGTCCCACC AGCTCCATGG		833
30	GAGTGACCAC CTCAGTCCAG GGAAAACAGC TGAAGACCAA GATGGTGAGC TCTGATTGCT TCAGTTGGTC ATCAACTATT TTCCCTTGAC TGCTGTCCTG GGATGCCGG CTATTTGAT GGATAGATTG TGAATATCAG GAGGCCAGGG ATCACTGTGG ACCATCTTAG CAGTTGACCT		953
35	AACACATCTT CTTTCAATA TCTAAGAACT TTTGCCACTG TGACTAATGG TCCTAATATT AAGCTGTTGT TTATATTTAT CATATATCTA TGGCTACATG GTTATATTAT GCTGTGGTTG CGTTCGGTTT TATTACAGT TGCTTTACA AATATTGCT GTAACATTG ACTTCTAAGG		1013
40	TTTAGATGCC ATTTAAGAAC TGAGATGGAT AGCTTTAAA GCATCTTTA CTTCTTACCA TTTTTAAAAA GTATGCAGCT AAATCGAAG CTTTGGTCT ATATTGTTAA TTGCCATTGC		1373
45	TGTAAATCTT AAAATGAATG AATAAAATG TTTCATTTA AAAAAAAAAA AAAAAAAAAA		1433
	AAAAA		1493
			1553
			1557

## 45 (2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 171 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

50	Gln Ser Leu Ser Thr Ser Gln Thr Phe Tyr Asp Ala Tyr Ile Ser Tyr		
	1	5	10
55			15
60	Asp Thr Lys Asp Ala Ser Val Thr Asp Trp Val Ile Asn Glu Leu Arg		
	20	25	30

Tyr His Leu Glu Glu Ser Arg Asp Lys Asn Val Leu Leu Cys Leu Glu  
 35 40 45

5 Glu Arg Asp Trp Asp Pro Gly Leu Ala Ile Ile Asp Asn Leu Met Gln  
 50 55 60

Ser Ile Asn Gln Ser Lys Lys Thr Val Phe Val Leu Thr Lys Lys Tyr  
 65 70 75 80

10 Ala Lys Ser Trp Asn Phe Lys Thr Ala Phe Tyr Leu Gly Leu Gln Arg.  
 85 90 95

15 Leu Met Gly Glu Asn Met Asp Val Ile Ile Phe Ile Leu Leu Glu Pro  
 100 105 110

Val Leu Gln His Ser Pro Tyr Leu Arg Leu Arg Gln Arg Ile Cys Lys  
 115 120 125

20 Ser Ser Ile Leu Gln Trp Pro Asp Asn Pro Lys Ala Glu Arg Leu Phe  
 130 135 140

Trp Gln Thr Leu Arg Asn Val Val Leu Thr Glu Asn Asp Ser Arg Tyr  
 145 150 155 160

25 Asn Asn Met Tyr Val Asp Ser Ile Lys Gln Tyr  
 165 170

## (2) INFORMATION FOR SEQ ID NO:19:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 629 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: cDNA

40 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..486

45 (ix) FEATURE:  
 (A) NAME/KEY: misc\_feature  
 (B) LOCATION: 144  
 (D) OTHER INFORMATION: /note= "nucleotides 144 and 225  
 designated C; may be C or T"

## 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAT GAA TTG ATC CCC AAT CTA GAG AAG GAA GAT GGT TCT ATC TTG ATT	48
Asn Glu Leu Ile Pro Asn Leu Glu Lys Glu Asp Gly Ser Ile Leu Ile	
1 5 10 15	
55 TGC CTT TAT GAA AGC TAC TTT GAC CCT GGC AAA AGC ATT AGT GAA AAT	96
Cys Leu Tyr Glu Ser Tyr Phe Asp Pro Gly Lys Ser Ile Ser Glu Asn	
20 25 30	
60 ATT GTA AGC TTC ATT GAG AAA AGC TAT AAG TCC ATC TTT GTT TTG TCC	144
Ile Val Ser Phe Ile Glu Lys Ser Tyr Lys Ser Ile Phe Val Leu Ser	

	35	40	45	
	CCC AAC TTT GTC CAG AAT GAG TGG TGC CAT TAT GAA TTC TAC TTT GCC			192
5	Pro Asn Phe Val Gln Asn Glu Trp Cys His Tyr Glu Phe Tyr Phe Ala			
	50	55	60	
	CAC CAC AAT CTC TTC CAT GAA AAT TCT GAT CAC ATA ATT CTT ATC TTA			240
	His His Asn Leu Phe His Glu Asn Ser Asp His Ile Ile Leu Ile Leu			
10	65	70	75	80
	CTG GAA CCC ATT CCA TTC TAT TGC ATT CCC ACC AGG TAT CAT AAA CTG			288
	Leu Glu Pro Ile Pro Phe Tyr Cys Ile Pro Thr Arg Tyr His Lys Leu			
	85	90	95	
15	GAA GCT CTC CTG GAA AAA AAA GCA TAC TTG GAA TGG CCC AAG GAT AGG			336
	Glu Ala Leu Leu Glu Lys Lys Ala Tyr Leu Glu Trp Pro Lys Asp Arg			
	100	105	110	
20	CGT AAA TGT GGG CTT TTC TGG GCA AAC CTT CGA GCT GCT GTT AAT GTT			384
	Arg Lys Cys Gly Leu Phe Trp Ala Asn Leu Arg Ala Ala Val Asn Val			
	115	120	125	
25	AAT GTA TTA GCC ACC AGA GAA ATG TAT GAA CTG CAG ACA TTC ACA GAG			432
	Asn Val Leu Ala Thr Arg Glu Met Tyr Glu Leu Gln Thr Phe Thr Glu			
	130	135	140	
	TTA AAT GAA GAG TCT CGA GGT TCT ACA ATC TCT CTG ATG AGA ACA GAC			480
	Leu Asn Glu Glu Ser Arg Gly Ser Thr Ile Ser Leu Met Arg Thr Asp			
30	145	150	155	160
	TGT CTA TAAAATCCCA CAGTCCTTGG GAAGTTGGGG ACCACATACA CTGTTGGGAT			536
	Cys Leu			
35	GTACATTGAT ACAACCTTA TGATGGCAAT TTGACAATAT TTATTAAAAT AAAAAATGGT			596
	TATTCCCTTC AAAAAAAAAA AAAAAAAAAA AAA			629

## 40 (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 amino acids
- (B) TYPE: amino acid
- 45 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

	Asn Glu Leu Ile Pro Asn Leu Glu Lys Glu Asp Gly Ser Ile Leu Ile			
50	1	5	10	15
	Cys Leu Tyr Glu Ser Tyr Phe Asp Pro Gly Lys Ser Ile Ser Glu Asn			
55	20	25	30	
	Ile Val Ser Phe Ile Glu Lys Ser Tyr Lys Ser Ile Phe Val Leu Ser			
	35	40	45	
60	Pro Asn Phe Val Gln Asn Glu Trp Cys His Tyr Glu Phe Tyr Phe Ala			
	50	55	60	

	His His Asn Leu Phe His Glu Asn Ser Asp His Ile Ile Leu Ile Leu			
	65	70	75	80
5	Leu Glu Pro Ile Pro Phe Tyr Cys Ile Pro Thr Arg Tyr His Lys Leu			
	85	90	95	
	Glu Ala Leu Leu Glu Lys Lys Ala Tyr Leu Glu Trp Pro Lys Asp Arg			
10	100	105	110	
	Arg Lys Cys Gly Leu Phe Trp Ala Asn Leu Arg Ala Ala Val Asn Val			
	115	120	125	
15	Asn Val Leu Ala Thr Arg Glu Met Tyr Glu Leu Gln Thr Phe Thr Glu			
	130	135	140	
	Leu Asn Glu Glu Ser Arg Gly Ser Thr Ile Ser Leu Met Arg Thr Asp			
	145	150	155	160
20	Cys Leu			

## (2) INFORMATION FOR SEQ ID NO:21:

25           (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 427 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30           (ii) MOLECULE TYPE: cDNA

35           (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..426

## 40           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

40	AAG AAC TCC AAA GAA AAC CTC CAG TTT CAT GCT TTT ATT TCA TAT AGT Lys Asn Ser Lys Glu Asn Leu Gln Phe His Ala Phe Ile Ser Tyr Ser	48
	1               5               10               15	
45	GAA CAT GAT TCT GCC TGG GTG AAA AGT GAA TTG GTA CCT TAC CTA GAA Glu His Asp Ser Ala Trp Val Lys Ser Glu Leu Val Pro Tyr Leu Glu	96
	20               25               30	
50	AAA GAA GAT ATA CAG ATT TGT CTT CAT GAG AGA AAC TTT GTC CCT GGC Lys Glu Asp Ile Gln Ile Cys Leu His Glu Arg Asn Phe Val Pro Gly	144
	35               40               45	
55	AAG AGC ATT GTG GAA AAT ATC ATC AAC TGC ATT GAG AAG AGT TAC AAG Lys Ser Ile Val Glu Asn Ile Ile Asn Cys Ile Glu Lys Ser Tyr Lys	192
	50               55               60	
60	TCC ATC TTT GTT TTG TCT CCC AAC TTT GTC CAG AGT GAG TGG TGC CAT Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Ser Glu Trp Cys His	240
	65               70               75               80	
	TAC GAA CTC TAT TTT GCC CAT CAC AAT CTC TTT CAT GAA GGA TCT AAT	288

	Tyr Glu Leu Tyr Phe Ala His His Asn Leu Phe His Glu Gly Ser Asn		
	85	90	95
5	AAC TTA ATC CTC ATC TTA CTG GAA CCC ATT CCA CAG AAC AGC ATT CCC Asn Leu Ile Leu Ile Leu Leu Glu Pro Ile Pro Gln Asn Ser Ile Pro		336
	100	105	110
10	AAC AAG TAC CAC AAG CTG AAG GCT CTC ATG ACG CAG CGG ACT TAT TTG Asn Lys Tyr His Lys Leu Lys Ala Leu Met Thr Gln Arg Thr Tyr Leu		384
	115	120	125
15	CAG TGG CCC AAG GAG AAA AGC AAA CGT GGG CTC TTT TGG GCT Gln Trp Pro Lys Glu Lys Ser Lys Arg Gly Leu Phe Trp Ala		426
	130	135	140
	A		427
20	(2) INFORMATION FOR SEQ ID NO:22:		
	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 142 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
25	(ii) MOLECULE TYPE: protein		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:		
30	Lys Asn Ser Lys Glu Asn Leu Gln Phe His Ala Phe Ile Ser Tyr Ser		
	1	5	10
	15		
	Glu His Asp Ser Ala Trp Val Lys Ser Glu Leu Val Pro Tyr Leu Glu		
	20	25	30
35	Lys Glu Asp Ile Gln Ile Cys Leu His Glu Arg Asn Phe Val Pro Gly		
	35	40	45
40	Lys Ser Ile Val Glu Asn Ile Ile Asn Cys Ile Glu Lys Ser Tyr Lys		
	50	55	60
	Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Ser Glu Trp Cys His		
	65	70	75
	80		
45	Tyr Glu Leu Tyr Phe Ala His His Asn Leu Phe His Glu Gly Ser Asn		
	85	90	95
	Asn Leu Ile Leu Ile Leu Glu Pro Ile Pro Gln Asn Ser Ile Pro		
50	100	105	110
	Asn Lys Tyr His Lys Leu Lys Ala Leu Met Thr Gln Arg Thr Tyr Leu		
	115	120	125
55	Gln Trp Pro Lys Glu Lys Ser Lys Arg Gly Leu Phe Trp Ala		
	130	135	140
	(2) INFORMATION FOR SEQ ID NO:23:		
60	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 662 base pairs		
	(B) TYPE: nucleic acid		

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..627

10

(ix) FEATURE:

(A) NAME/KEY: misc\_feature  
(B) LOCATION: 54

15

(D) OTHER INFORMATION: /note= "nucleotides 54, 103, and 345 are designated A; each may be A or G"

20

(ix) FEATURE:

(A) NAME/KEY: misc\_feature  
(B) LOCATION: 313

(D) OTHER INFORMATION: /note= "nucleotide 313 designated G, may be G or T"

25

(ix) FEATURE:

(A) NAME/KEY: misc\_feature  
(B) LOCATION: 316

(D) OTHER INFORMATION: /note= "nucleotides 316, 380, 407, and 408 designated C; each may be A, C, G, or T"

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCT TCC ACC TGT GCC TGG CCT GGC TTC CCT GGC GGG GGC GGC AAA GTG  
Ala Ser Thr Cys Ala Trp Pro Gly Phe Pro Gly Gly Gly Lys Val  
1 5 10 15

48

35

GGC GAA ATG AGG ATG CCC TGC CCT ACG ATG CCT TCG TGG TCT TCG ACA  
Gly Glu Met Arg Met Pro Cys Pro Thr Met Pro Ser Trp Ser Ser Thr  
20 25 30

96

40

AAA CGC AGA GCG CAG TGG CAG ACT GGG TGT ACA ACG AGC TTC GGG GGC  
Lys Arg Arg Ala Gln Trp Gln Thr Gly Cys Thr Thr Ser Phe Gly Gly  
35 40 45

144

45

AGC TGG AGG AGT GCC GTG GGC GCT GGG CAC TCC GCC TGT GCC TGG AGG  
Ser Trp Arg Ser Ala Val Gly Ala Gly His Ser Ala Cys Ala Trp Arg  
50 55 60

192

50

AAC GCG ACT GGC TGC CTG GCA AAA CCC TCT TTG AGA ACC TGT GGG CCT  
Asn Ala Thr Gly Cys Leu Ala Lys Pro Ser Leu Arg Thr Cys Gly Pro  
65 70 75 80

240

55

CGG TCT ATG GCA GCC GCA AGA CGC TGT TTG TGC TGG CCC ACA CGG ACC  
Arg Ser Met Ala Ala Ala Arg Arg Cys Leu Cys Trp Pro Thr Arg Thr  
85 90 95

288

60

GGG TCA GTG GTC TCT TGC GCG CCA GTT CTC CTG CTG GCC CAG CAG CGC  
Gly Ser Val Val Ser Cys Ala Pro Val Leu Leu Leu Ala Gln Gln Arg  
100 105 110

336

CTG CTG GAA GAC CGC AAG GAC GTC GTG GTG CTG GTG ATC CTA ACG CCT  
Leu Leu Glu Asp Arg Lys Asp Val Val Leu Val Ile Leu Thr Pro

384

	115	120	125	
	GAC GGC CAA GCC TCC CGA CTA CCC GAT GCG CTG ACC AGC GCC TCT GCC			432
5	Asp Gly Gln Ala Ser Arg Leu Pro Asp Ala Leu Thr Ser Ala Ser Ala			
	130	135	140	
	GCC AGA GTG TCC TCC TCT GGC CCC ACC AGC CCA GTG GTC GCG CAG CTT			480
	Ala Arg Val Ser Ser Gly Pro Thr Ser Pro Val Val Ala Gln Leu			
10	145	150	155	160
	CTG AGG CCA GCA TGC ATG GCC CTG ACC AGG GAC AAC CAC CAC TTC TAT			528
	Leu Arg Pro Ala Cys Met Ala Leu Thr Arg Asp Asn His His Phe Tyr			
	165	170	175	
15	AAC CGG AAC TTC TGC CAG GGA ACC CAC GGC CGA ATA GCC GTG AGC CGG			576
	Asn Arg Asn Phe Cys Gln Gly Thr His Gly Arg Ile Ala Val Ser Arg			
	180	185	190	
20	AAT CCT GCA CGG TGC CAC CTC CAC ACA CAC CTA ACA TAT GCC TGC CTG			624
	Asn Pro Ala Arg Cys His Leu His Thr His Leu Thr Tyr Ala Cys Leu			
	195	200	205	
25	ATC TGACCAACAC ATGCTGCCA CCCTCACCCAC ACACC Ile			662

## (2) INFORMATION FOR SEQ ID NO:24:

30           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 209 amino acids  
               (B) TYPE: amino acid  
               (D) TOPOLOGY: linear

35           (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

	Ala Ser Thr Cys Ala Trp Pro Gly Phe Pro Gly Gly Gly Lys Val			
40	1	5	10	15
	Gly Glu Met Arg Met Pro Cys Pro Thr Met Pro Ser Trp Ser Ser Thr			
	20	25	30	
45	Lys Arg Arg Ala Gln Trp Gln Thr Gly Cys Thr Thr Ser Phe Gly Gly			
	35	40	45	
	Ser Trp Arg Ser Ala Val Gly Ala Gly His Ser Ala Cys Ala Trp Arg			
50	50	55	60	
	Asn Ala Thr Gly Cys Leu Ala Lys Pro Ser Leu Arg Thr Cys Gly Pro			
	65	70	75	80
55	Arg Ser Met Ala Ala Arg Arg Cys Leu Cys Trp Pro Thr Arg Thr			
	85	90	95	
	Gly Ser Val Val Ser Cys Ala Pro Val Leu Leu Ala Gln Gln Arg			
	100	105	110	
60	Leu Leu Glu Asp Arg Lys Asp Val Val Val Leu Val Ile Leu Thr Pro			
	115	120	125	

Asp Gly Gln Ala Ser Arg Leu Pro Asp Ala Leu Thr Ser Ala Ser Ala  
130 135 140

5 Ala Arg Val Ser Ser Ser Gly Pro Thr Ser Pro Val Val Ala Gln Leu  
145 150 155 160

Leu Arg Pro Ala Cys Met Ala Leu Thr Arg Asp Asn His His Phe Tyr  
165 170 175

10 Asn Arg Asn Phe Cys Gln Gly Thr His Gly Arg Ile Ala Val Ser Arg  
180 185 190

Asn Pro Ala Arg Cys His Leu His Thr His Leu Thr Tyr Ala Cys Leu  
15 195 200 205

Ile

20 (2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 107..2617

35

(ix) FEATURE:

- (A) NAME/KEY: mat\_peptide
- (B) LOCATION: 173..2617

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 81
- (D) OTHER INFORMATION: /note= "nucleotides 81, 3144, 3205, and 3563 designated A, each may be A, C, G, or T"

45

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 84
- (D) OTHER INFORMATION: /note= "nucleotide 84 designated C, may be C or G"

50

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 739
- (D) OTHER INFORMATION: /note= "nucleotide 739 designated C, may be C or T"

55

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 3132
- (D) OTHER INFORMATION: /note= "nucleotides 3132, 3532, 3538, and 3553 designated G, each may be G or T"

## (ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION: 3638

5 (D) OTHER INFORMATION: /note= "nucleotide 3638 designated A, may be A or T"

## (ix) FEATURE:

(A) NAME/KEY: misc\_feature

10 (B) LOCATION: 3677

(D) OTHER INFORMATION: /note= "nucleotides 3677, 3685, and 3736 designated C, each may be A or C"

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

	AAAATACTCC CTTGCCTCAA AAACTGCTCG GTCAACGGT GATAGCAAAC CACGCATTCA	60
20	CAGGGCCACT GCTGCTCACA AAACCAGTGA GGATGATGCC AGGATG ATG TCT GCC Met Ser Ala -22 -20	115
25	TCG CGC CTG GCT GGG ACT CTG ATC CCA GCC ATG GCC TTC CTC TCC TGC Ser Arg Leu Ala Gly Thr Leu Ile Pro Ala Met Ala Phe Leu Ser Cys -15 -10 -5	163
30	GTG AGA CCA GAA AGC TGG GAG CCC TGC GTG GAG GTT CCT AAT ATT ACT Val Arg Pro Glu Ser Trp Glu Pro Cys Val Glu Val Pro Asn Ile Thr 1 5 10	211
35	TAT CAA TGC ATG GAG CTG AAT TTC TAC AAA ATC CCC GAC AAC CTC CCC Tyr Gln Cys Met Glu Leu Asn Phe Tyr Lys Ile Pro Asp Asn Leu Pro 15 20 25	259
40	TTC TCA ACC AAG AAC CTG GAC CTG AGC TTT AAT CCC CTG AGG CAT TTA Phe Ser Thr Lys Asn Leu Asp Leu Ser Phe Asn Pro Leu Arg His Leu 30 35 40 45	307
45	GGC AGC TAT AGC TTC AGT TTC CCA GAA CTG CAG GTG CTG GAT TTA Gly Ser Tyr Ser Phe Phe Ser Phe Pro Glu Leu Gln Val Leu Asp Leu 50 55 60	355
50	TCC AGG TGT GAA ATC CAG ACA ATT GAA GAT GGG GCA TAT CAG AGC CTA Ser Arg Cys Glu Ile Gln Thr Ile Glu Asp Gly Ala Tyr Gln Ser Leu 65 70 75	403
55	AGC CAC CTC TCT ACC TTA ATA TTG ACA GGA AAC CCC ATC CAG AGT TTA Ser His Leu Ser Thr Leu Ile Leu Thr Gly Asn Pro Ile Gln Ser Leu 80 85 90	451
60	GCC CTG GGA GCC TTT TCT GGA CTA TCA AGT TTA CAG AAG CTG GTG GCT Ala Leu Gly Ala Phe Ser Gly Leu Ser Ser Leu Gln Lys Leu Val Ala 95 100 105	499
65	GTG GAG ACA AAT CTA GCA TCT CTA GAG AAC TTC CCC ATT GGA CAT CTC Val Glu Thr Asn Leu Ala Ser Leu Glu Asn Phe Pro Ile Gly His Leu 110 115 120 125	547
70	AAA ACT TTG AAA GAA CTT AAT GTG GCT CAC AAT CTT ATC CAA TCT TTC Lys Thr Leu Lys Glu Leu Asn Val Ala His Asn Leu Ile Gln Ser Phe 130 135 140	595

	AAA TTA CCT GAG TAT TTT TCT AAT CTG ACC AAT CTA GAG CAC TTG GAC Lys Leu Pro Glu Tyr Phe Ser Asn Leu Thr Asn Leu Glu His Leu Asp 145 150 155	643
5	CTT TCC AGC AAC AAG ATT CAA AGT ATT TAT TGC ACA GAC TTG CGG GTT Leu Ser Ser Asn Lys Ile Gln Ser Ile Tyr Cys Thr Asp Leu Arg Val 160 165 170	691
10	CTA CAT CAA ATG CCC CTA CTC AAT CTC TCT TTA GAC CTG TCC CTG AAC. Leu His Gln Met Pro Leu Leu Asn Leu Ser Leu Asp Leu Ser Leu Asn 175 180 185	739
15	CCT ATG AAC TTT ATC CAA CCA GGT GCA TTT AAA GAA ATT AGG CTT CAT Pro Met Asn Phe Ile Gln Pro Gly Ala Phe Lys Glu Ile Arg Leu His 190 195 200 205	787
20	AAG CTG ACT TTA AGA AAT AAT TTT GAT AGT TTA AAT GTA ATG AAA ACT Lys Leu Thr Leu Arg Asn Asn Phe Asp Ser Leu Asn Val Met Lys Thr 210 215 220	835
25	TGT ATT CAA GGT CTG GCT GGT TTA GAA GTC CAT CGT TTG GTT CTG GGA Cys Ile Gln Gly Leu Ala Gly Leu Glu Val His Arg Leu Val Leu Gly 225 230 235	883
30	GAA TTT AGA AAT GAA GGA AAC TTG GAA AAG TTT GAC AAA TCT GCT CTA Glu Phe Arg Asn Glu Gly Asn Leu Glu Lys Phe Asp Lys Ser Ala Leu 240 245 250	931
35	GAG GGC CTG TGC AAT TTG ACC ATT GAA GAA TTC CGA TTA GCA TAC TTA Glu Gly Leu Cys Asn Leu Thr Ile Glu Glu Phe Arg Leu Ala Tyr Leu 255 260 265	979
40	GAC TAC TAC CTC GAT GAT ATT ATT GAC TTA TTT AAT TGT TTG ACA AAT Asp Tyr Tyr Leu Asp Asp Ile Ile Asp Leu Phe Asn Cys Leu Thr Asn 270 275 280 285	1027
45	GTT TCT TCA TTT TCC CTG GTG AGT GTG ACT ATT GAA AGG GTA AAA GAC Val Ser Ser Phe Ser Leu Val Ser Val Thr Ile Glu Arg Val Lys Asp 290 295 300	1075
50	TTT TCT TAT AAT TTC GGA TGG CAA CAT TTA GAA TTA GTT AAC TGT AAA Phe Ser Tyr Asn Phe Gly Trp Gln His Leu Glu Leu Val Asn Cys Lys 305 310 315	1123
55	TTT GGA CAG TTT CCC ACA TTG AAA CTC AAA TCT CTC AAA AGG CTT ACT Phe Gly Gln Phe Pro Thr Leu Lys Leu Lys Ser Leu Lys Arg Leu Thr 320 325 330	1171
60	TTC ACT TCC AAC AAA GGT GGG AAT GCT TTT TCA GAA GTT GAT CTA CCA Phe Thr Ser Asn Lys Gly Gly Asn Ala Phe Ser Glu Val Asp Leu Pro 335 340 345	1219
	AGC CTT GAG TTT CTA GAT CTC AGT AGA AAT GGC TTG AGT TTC AAA GGT Ser Leu Glu Phe Leu Asp Leu Ser Arg Asn Gly Leu Ser Phe Lys Gly 350 355 360 365	1267
	TGC TGT TCT CAA AGT GAT TTT GGG ACA ACC AGC CTA AAG TAT TTA GAT Cys Cys Ser Gln Ser Asp Phe Gly Thr Thr Ser Leu Lys Tyr Leu Asp 370 375 380	1315

	CTG AGC TTC AAT GGT GTT ATT ACC ATG AGT TCA AAC TTC TTG GGC TTA		1363
	Leu Ser Phe Asn Gly Val Ile Thr Met Ser Ser Asn Phe Leu Gly Leu		
	385 390 395		
5	GAA CAA CTA GAA CAT CTG GAT TTC CAG CAT TCC AAT TTG AAA CAA ATG		1411
	Glu Gln Leu Glu His Leu Asp Phe Gln His Ser Asn Leu Lys Gln Met		
	400 405 410		
10	AGT GAG TTT TCA GTA TTC CTA TCA CTC AGA AAC CTC ATT TAC CTT GAC		1459
	Ser Glu Phe Ser Val Phe Leu Ser Leu Arg Asn Leu Ile Tyr Leu Asp		
	415 420 425		
15	ATT TCT CAT ACT CAC ACC AGA GTT GCT TTC AAT GGC ATC TTC AAT GGC		1507
	Ile Ser His Thr His Thr Arg Val Ala Phe Asn Gly Ile Phe Asn Gly		
	430 435 440 445		
20	TTG TCC AGT CTC GAA GTC TTG AAA ATG GCT GGC AAT TCT TTC CAG GAA		1555
	Leu Ser Ser Leu Glu Val Leu Lys Met Ala Gly Asn Ser Phe Gln Glu		
	450 455 460		
25	AAC TTC CTT CCA GAT ATC TTC ACA GAG CTG AGA AAC TTG ACC TTC CTG		1603
	Asn Phe Leu Pro Asp Ile Phe Thr Glu Leu Arg Asn Leu Thr Phe Leu		
	465 470 475		
30	GAC CTC TCT CAG TGT CAA CTG GAG CAG TTG TCT CCA ACA GCA TTT AAC		1651
	Asp Leu Ser Gln Cys Gln Leu Glu Gln Leu Ser Pro Thr Ala Phe Asn		
	480 485 490		
35	TCA CTC TCC AGT CTT CAG GTA CTA AAT ATG AGC CAC AAC AAC TTC TTT		1699
	Ser Leu Ser Ser Leu Gln Val Leu Asn Met Ser His Asn Asn Phe Phe		
	495 500 505		
40	TCA TTG GAT ACG TTT CCT TAT AAG TGT CTG AAC TCC CTC CAG GTT CTT		1747
	Ser Leu Asp Thr Phe Pro Tyr Lys Cys Leu Asn Ser Leu Gln Val Leu		
	510 515 520 525		
45	GAT TAC AGT CTC AAT CAC ATA ATG ACT TCC AAA AAA CAG GAA CTA CAG		1795
	Asp Tyr Ser Leu Asn His Ile Met Thr Ser Lys Lys Gln Glu Leu Gln		
	530 535 540		
50	CAT TTT CCA AGT AGT CTA GCT TTC TTA AAT CTT ACT CAG AAT GAC TTT		1843
	His Phe Pro Ser Ser Leu Ala Phe Leu Asn Leu Thr Gln Asn Asp Phe		
	545 550 555		
55	GCT TGT ACT TGT GAA CAC CAG AGT TTC CTG CAA TGG ATC AAG GAC CAG		1891
	Ala Cys Thr Cys Glu His Gln Ser Phe Leu Gln Trp Ile Lys Asp Gln		
	560 565 570		
60	AGG CAG CTC TTG GTG GAA GTT GAA CGA ATG GAA TGT GCA ACA CCT TCA		1939
	Arg Gln Leu Leu Val Glu Val Glu Arg Met Glu Cys Ala Thr Pro Ser		
	575 580 585		
65	GAT AAG CAG GGC ATG CCT GTG CTG AGT TTG AAT ATC ACC TGT CAG ATG		1987
	Asp Lys Gln Gly Met Pro Val Leu Ser Leu Asn Ile Thr Cys Gln Met		
	590 595 600 605		
70	AAT AAG ACC ATC ATT GGT GTG TCG GTC CTC AGT GTG CTT GTA GTA TCT		2035
	Asn Lys Thr Ile Ile Gly Val Ser Val Leu Ser Val Leu Val Val Ser		
	610 615 620		
75	GTT GTA GCA GTT CTG GTC TAT AAG TTC TAT TTT CAC CTG ATG CTT CTT		2083

	Val Val Ala Val Leu Val Tyr Lys Phe Tyr Phe His Leu Met Leu Leu	
	625 . . . . . 630 . . . . . 635	
5	GCT GGC TGC ATA AAG TAT GGT AGA GGT GAA AAC ATC TAT GAT GCC TTT Ala Gly Cys Ile Lys Tyr Gly Arg Gly Glu Asn Ile Tyr Asp Ala Phe 640 . . . . . 645 . . . . . 650	2131
10	GTT ATC TAC TCA AGC CAG GAT GAG GAC TGG GTA AGG AAT GAG CTA GTA Val Ile Tyr Ser Ser Gln Asp Glu Asp Trp Val Arg Asn Glu Leu Val 655 . . . . . 660 . . . . . 665	2179
15	AAG AAT TTA GAA GAA GGG GTG CCT CCA TTT CAG CTC TGC CTT CAC TAC Lys Asn Leu Glu Glu Gly Val Pro Pro Phe Gln Leu Cys Leu His Tyr 670 . . . . . 675 . . . . . 680 . . . . . 685	2227
20	AGA GAC TTT ATT CCC GGT GTG GCC ATT GCT GCC AAC ATC ATC CAT GAA Arg Asp Phe Ile Pro Gly Val Ala Ile Ala Ala Asn Ile Ile His Glu 690 . . . . . 695 . . . . . 700	2275
25	GGT TTC CAT AAA AGC CGA AAG GTG ATT GTT GTG GTG TCC CAG CAC TTC Gly Phe His Lys Ser Arg Lys Val Ile Val Val Val Ser Gln His Phe 705 . . . . . 710 . . . . . 715	2323
30	ATC CAG AGC CGC TGG TGT ATC TTT GAA TAT GAG ATT GCT CAG ACC TGG Ile Gln Ser Arg Trp Cys Ile Phe Glu Tyr Glu Ile Ala Gln Thr Trp 720 . . . . . 725 . . . . . 730	2371
35	CAG TTT CTG AGC AGT CGT GCT GGT ATC ATC TTC ATT GTC CTG CAG AAG Gln Phe Leu Ser Ser Arg Ala Gly Ile Ile Phe Ile Val Leu Gln Lys 735 . . . . . 740 . . . . . 745	2419
40	GTG GAG AAG ACC CTG CTC AGG CAG CAG GTG GAG CTG TAC CGC CTT CTC Val Glu Lys Thr Leu Leu Arg Gln Gln Val Glu Leu Tyr Arg Leu Leu 750 . . . . . 755 . . . . . 760 . . . . . 765	2467
45	AGC AGG AAC ACT TAC CTG GAG TGG GAG GAC AGT GTC CTG GGG CGG CAC Ser Arg Asn Thr Tyr Leu Glu Trp Glu Asp Ser Val Leu Gly Arg His 770 . . . . . 775 . . . . . 780	2515
50	ATC TTC TGG AGA CGA CTC AGA AAA GCC CTG CTG GAT GGT AAA TCA TGG Ile Phe Trp Arg Arg Leu Arg Lys Ala Leu Leu Asp Gly Lys Ser Trp 785 . . . . . 790 . . . . . 795	2563
55	AAT CCA GAA GGA ACA GTG GGT ACA GGA TGC AAT TGG CAG GAA GCA ACA Asn Pro Glu Gly Thr Val Gly Thr Gly Cys Asn Trp Gln Glu Ala Thr 800 . . . . . 805 . . . . . 810	2611
60	TCT ATC TGAAGAGGAA AAATAAAAAC CTCCTGAGGC ATTTCTTGCC CAGCTGGTCA Ser Ile 815	2667
	CAACACTTGT TCAGTTAATA AGTATTAAAT GCTGCCACAT GTCAGGCCTT ATGCTAAGGG	2727
	TGAGTAATTG CATGGTGCAC TAGATATGCA GGGCTGCTAA TCTCAAGGAG CTTCCAGTGC	2787
	AGAGGGAAATA AATGCTAGAC TAAAATACAG AGTCTTCCAG GTGGGCATTT CAACCAAAC	2847
	AGTCAAGGAA CCCATGACAA AGAAAGTCAT TTCAACTCTT ACCTCATCAA GTTGAATAAA	2907
	GACAGAGAAA ACAGAAAGAG ACATTGTTCT TTTCTGAGT CTTTTGAATG GAAATTGTAT	2967

	TATGTTATAG CCATCATAAA ACCATTTGG TAGTTTGAC TGAACGGGT GTTCACTTT	3027
	TCCTTTGA TTGAATACAA TTTAAATTCT ACTTGATGAC TGCAGTCGTC AAGGGGCTCC	3087
5	TGATGCAAGA TGCCCCTCC ATTAAAGTC TGTCTCCTTA CAGAGGTTAA AGTCTAATGG	3147
	CTAATTCCCTA AGGAAACCTG ATTAACACAT GCTCACACC ATCCTGGTCA TTCTCGAAC	3207
10	TGTTCTATTT TTTAACTAAT CACCCCTGAT ATATTTTAT TTTTATATAT CCAGTTTCA	3267
	TTTTTTACG TCTTGCCTAT AAGCTAATAT CATAAATAAG GTTGTAAAG ACgtGCTTCA	3327
	AATATCCATA TTAACCCTA TTTTCAAGG AAGTATGGAA AAGTACACTC TGTCACTTG	3387
15	TCACTCGATG TCATTCCAAA GTTATTGCCT ACTAAGTAAT GACTGTCATG AAAGCAGCAT	3447
	TGAAATAATT TGTTAAAGG GGGCACTCTT TTAAACGGGA AGAAAATTTC CGCTTCCTGG	3507
20	TCTTATCATG GACAATTGG GCTAGAGGCA GGAAGGAAGT GGGATGACCT CAGGAAGTCA	3567
	CCTTTCTTG ATTCCAGAAA CATATGGGCT GATAAACCG GGGTGACCTC ATGAAATGAG	3627
	TTGCAGCAGA AGTTTATTT TTTCAGAACCA AGTGTGTTT GATGGACCTC TGAATCTCTT	3687
25	TAGGGAGACA CAGATGGCTG GGATCCCTCC CCTGTACCCCT TCTCACTGCC AGGAGAACTA	3747
	CGTGTGAAGG TATTCAAGGC AGGGAGTATA CATTGCTGTT TCCTGTTGGG CAATGCTCCT	3807
30	TGACCACATT TTGGGAAGAG TGGATGTTAT CATTGAGAAA ACAATGTGTC TGGAATTAAT	3867
	GGGTTCTTA TAAAGAAGGT TCCCAGAAAA GAATGTCAT TCCAGCTTCT TCAGGAAACA	3927
	GGAACATTCA AGGAAAAGGA CAATCAGGAT GTCATCAGGG AAATGAAAAT AAAAACACACA	3987
35	ATGAGATATC ACCTTATACC AGGTAGATGG CTACTATAAA AAAATGAAGT GTCACTCAAGG	4047
	ATATAGAGAA ATTGGAACCC TTCTTCACTG CTGGAGGGAA TGGAAAATGG TGTAGCCGTT	4107
40	ATGAAAACA GTACGGAGGT TTCTAAAAA TTAAAAATAG AACTGCTATA TGATCCAGCA	4167
	ATCTCACTTC TGTATATATA CCCAAAATAA TTGAAATCAG AATTCAAGA AAATATTTAC	4227
	ACTCCCATGT TCATTGTGGC ACTCTTCACA ATCACTGTTT CCAAAGTTAT GGAAACAACC	4287
45	CAAATTCCA TTGGAAAATA AATGGACAAA GGAAATGTGC ATATAACGTA CAATGGGGAT	4347
	ATTATTCAGC CTAAAAAAAG GGGGGATCCT GTTATTATG ACAACATGAA TAAACCCGGA	4407
50	GGCCATTATG CTATGTAAA TGAGCAAGTA ACAGAAAGAC AAATACTGCC TGATTCATT	4467
	TATATGAGGT TCTAAAATAG TCAAACATCAT AGAACAGAG AATAGAACAG TGGTTCTTAG	4527
	GGAAAAGGAG GAAGGGAGAA ATGAGGAAAT AGGGAGTTGT CTAATTGGTA TAAAATTATA	4587
55	GTATGCAAGA TGAATTAGCT CTAAAGATCA GCTGTATAGC AGAGTCGTA TAATGAACAA	4647
	TACTGTATTA TGCACCTAAC ATTTGTTAA GAGGGTACCT CTCATGTTAA GTGTTCTTAC	4707
60	CATATACATA TACACAAGGA AGCTTTGGA GGTGATGGAT ATATTATTA CCTTGATTGT	4767
	GGTGATGGTT TGACAGGTAT GTGACTATGT CTAAACTCAT CAAATTGTAT ACATTAATA	4827

TATGCAGTTT TATAATATCA AAAAAAAA AAAAAAAA

4865

## 5 (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 837 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

10	Met Ser Ala Ser Arg Leu Ala Gly Thr Leu Ile Pro Ala Met Ala Phe	-22	-20	-15	-10
15	Leu Ser Cys Val Arg Pro Glu Ser Trp Glu Pro Cys Val Glu Val Pro	-5	1	5	10
20	Asn Ile Thr Tyr Gln Cys Met Glu Leu Asn Phe Tyr Lys Ile Pro Asp	15	20	25	
25	Asn Leu Pro Phe Ser Thr Lys Asn Leu Asp Leu Ser Phe Asn Pro Leu	30	35	40	
30	Arg His Leu Gly Ser Tyr Ser Phe Phe Ser Phe Pro Glu Leu Gln Val	45	50	55	
35	Leu Asp Leu Ser Arg Cys Glu Ile Gln Thr Ile Glu Asp Gly Ala Tyr	60	65	70	
40	Gln Ser Leu Ser His Leu Ser Thr Leu Ile Leu Thr Gly Asn Pro Ile	75	80	85	90
45	Gln Ser Leu Ala Leu Gly Ala Phe Ser Gly Leu Ser Ser Leu Gln Lys	95	100	105	
50	Leu Val Ala Val Glu Thr Asn Leu Ala Ser Leu Glu Asn Phe Pro Ile	110	115	120	
55	Gly His Leu Lys Thr Leu Lys Glu Leu Asn Val Ala His Asn Leu Ile	125	130	135	
60	Gln Ser Phe Lys Leu Pro Glu Tyr Phe Ser Asn Leu Thr Asn Leu Glu	140	145	150	
65	His Leu Asp Leu Ser Ser Asn Lys Ile Gln Ser Ile Tyr Cys Thr Asp	155	160	165	170
70	Leu Arg Val Leu His Gln Met Pro Leu Leu Asn Leu Ser Leu Asp Leu	175	180	185	
75	Ser Leu Asn Pro Met Asn Phe Ile Gln Pro Gly Ala Phe Lys Glu Ile	190	195	200	
80	Arg Leu His Lys Leu Thr Leu Arg Asn Asn Phe Asp Ser Leu Asn Val	205	210	215	
85	Met Lys Thr Cys Ile Gln Gly Leu Ala Gly Leu Glu Val His Arg Leu				

	220	225	230
	Val Leu Gly Glu Phe Arg Asn Glu Gly Asn Leu Glu Lys Phe Asp Lys		
5	235	240	245
	Ser Ala Leu Glu Gly Leu Cys Asn Leu Thr Ile Glu Glu Phe Arg Leu		
	255	260	265
10	Ala Tyr Leu Asp Tyr Tyr Leu Asp Asp Ile Ile Asp Leu Phe Asn Cys		
	270	275	280
	Leu Thr Asn Val Ser Ser Phe Ser Leu Val Ser Val Thr Ile Glu Arg		
	285	290	295
15	Val Lys Asp Phe Ser Tyr Asn Phe Gly Trp Gln His Leu Glu Leu Val		
	300	305	310
	Asn Cys Lys Phe Gly Gln Phe Pro Thr Leu Lys Leu Lys Ser Leu Lys		
20	315	320	325
	Arg Leu Thr Phe Thr Ser Asn Lys Gly Gly Asn Ala Phe Ser Glu Val		
	335	340	345
25	Asp Leu Pro Ser Leu Glu Phe Leu Asp Leu Ser Arg Asn Gly Leu Ser		
	350	355	360
	Phe Lys Gly Cys Cys Ser Gln Ser Asp Phe Gly Thr Thr Ser Leu Lys		
	365	370	375
30	Tyr Leu Asp Leu Ser Phe Asn Gly Val Ile Thr Met Ser Ser Asn Phe		
	380	385	390
	Leu Gly Leu Glu Gln Leu Glu His Leu Asp Phe Gln His Ser Asn Leu		
35	395	400	405
	Lys Gln Met Ser Glu Phe Ser Val Phe Leu Ser Leu Arg Asn Leu Ile		
	415	420	425
40	Tyr Leu Asp Ile Ser His Thr His Thr Arg Val Ala Phe Asn Gly Ile		
	430	435	440
	Phe Asn Gly Leu Ser Ser Leu Glu Val Leu Lys Met Ala Gly Asn Ser		
	445	450	455
45	Phe Gln Glu Asn Phe Leu Pro Asp Ile Phe Thr Glu Leu Arg Asn Leu		
	460	465	470
	Thr Phe Leu Asp Leu Ser Gln Cys Gln Leu Glu Gln Leu Ser Pro Thr		
50	475	480	485
	490		
	Ala Phe Asn Ser Leu Ser Ser Leu Gln Val Leu Asn Met Ser His Asn		
	495	500	505
55	Asn Phe Phe Ser Leu Asp Thr Phe Pro Tyr Lys Cys Leu Asn Ser Leu		
	510	515	520
	Gln Val Leu Asp Tyr Ser Leu Asn His Ile Met Thr Ser Lys Lys Gln		
	525	530	535
60	Glu Leu Gln His Phe Pro Ser Ser Leu Ala Phe Leu Asn Leu Thr Gln		
	540	545	550

	Asn Asp Phe Ala Cys Thr Cys Glu His Gln Ser Phe Leu Gln Trp Ile			
	555	560	565	570
5	Lys Asp Gln Arg Gln Leu Leu Val Glu Val Glu Arg Met Glu Cys Ala			
	575	580	585	
	Thr Pro Ser Asp Lys Gln Gly Met Pro Val Leu Ser Leu Asn Ile Thr			
	590	595	600	
10	Cys Gln Met Asn Lys Thr Ile Ile Gly Val Ser Val Leu Ser Val Leu			
	605	610	615	
15	Val Val Ser Val Val Ala Val Leu Val Tyr Lys Phe Tyr Phe His Leu			
	620	625	630	
	Met Leu Leu Ala Gly Cys Ile Lys Tyr Gly Arg Gly Glu Asn Ile Tyr			
	635	640	645	650
20	Asp Ala Phe Val Ile Tyr Ser Ser Gln Asp Glu Asp Trp Val Arg Asn			
	655	660	665	
	Glu Leu Val Lys Asn Leu Glu Glu Gly Val Pro Pro Phe Gln Leu Cys			
	670	675	680	
25	Leu His Tyr Arg Asp Phe Ile Pro Gly Val Ala Ile Ala Ala Asn Ile			
	685	690	695	
30	Ile His Glu Gly Phe His Lys Ser Arg Lys Val Ile Val Val Val Ser			
	700	705	710	
	Gln His Phe Ile Gln Ser Arg Trp Cys Ile Phe Glu Tyr Glu Ile Ala			
	715	720	725	730
35	Gln Thr Trp Gln Phe Leu Ser Ser Arg Ala Gly Ile Ile Phe Ile Val			
	735	740	745	
	Leu Gln Lys Val Glu Lys Thr Leu Leu Arg Gln Gln Val Glu Leu Tyr			
	750	755	760	
40	Arg Leu Leu Ser Arg Asn Thr Tyr Leu Glu Trp Glu Asp Ser Val Leu			
	765	770	775	
45	Gly Arg His Ile Phe Trp Arg Arg Leu Arg Lys Ala Leu Leu Asp Gly			
	780	785	790	
	Lys Ser Trp Asn Pro Glu Gly Thr Val Gly Thr Gly Cys Asn Trp Gln			
	795	800	805	810
50	Glu Ala Thr Ser Ile			
	815			

## (2) INFORMATION FOR SEQ ID NO:27:

- 55           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 300 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: single  
               (D) TOPOLOGY: linear
- 60           (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..300

(ix) FEATURE:  
(A) NAME/KEY: misc\_feature  
(B) LOCATION: 186  
(D) OTHER INFORMATION: /note= "nucleotides 186, 196, .217,  
276, and 300 designated C, each may be A, C, G, or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TCC	TAT	TCT	ATG	GAA	AAA	GAT	GCT	TTC	CTA	TTT	ATG	AGA	AAT	TTG	AAG		48
Ser	Tyr	Ser	Met	Glu	Lys	Asp	Ala	Phe	Leu	Phe	Met	Arg	Asn	Leu	Lys		
1			5						10					15			

GTT	CTC	TCA	CTA	AAA	GAT	AAC	AAT	GTC	ACA	GCT	GTC	CCC	ACC	ACT	TTG		96
Val	Leu	Ser	Leu	Lys	Asp	Asn	Asn	Val	Thr	Ala	Val	Pro	Thr	Thr	Leu		
	20					25						30					

CCA	CCT	AAT	TTA	CTA	GAG	CTC	TAT	CTT	TAT	AAC	AAT	ATC	ATT	AAG	AAA		144
Pro	Pro	Asn	Leu	Leu	Glu	Leu	Tyr	Leu	Tyr	Asn	Asn	Ile	Ile	Lys	Lys		
	35				40					45							

ATC	CAA	GAA	AAT	GAT	TTC	AAT	AAC	CTC	AAT	GAG	TTG	CAA	GTC	CTT	GAC		192
Ile	Gln	Glu	Asn	Asp	Phe	Asn	Asn	Leu	Asn	Glu	Leu	Gln	Val	Leu	Asp		
	50				55					60							

CTA	CGT	GGA	AAT	TGC	CCT	CGA	TGT	CAT	AAT	GTC	CCA	TAT	CCG	TGT	ACA		240
Leu	Arg	Gly	Asn	Cys	Pro	Arg	Cys	His	Asn	Val	Pro	Tyr	Pro	Cys	Thr		
	65				70				75			80					

CCG	TGT	GAA	AAT	AAT	TCC	CCC	TTA	CAG	ATC	CAT	GAC	AAT	GCT	TTC	AAT		288
Pro	Cys	Glu	Asn	Asn	Ser	Pro	Leu	Gln	Ile	His	Asp	Asn	Ala	Phe	Asn		
	85					90				95							

TCA	TCG	ACA	GAC													300	
Ser	Ser	Thr	Asp														
	100																

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 100 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ser	Tyr	Ser	Met	Glu	Lys	Asp	Ala	Phe	Leu	Phe	Met	Arg	Asn	Leu	Lys		50
1						5			10					15			

Val	Leu	Ser	Leu	Lys	Asp	Asn	Asn	Val	Thr	Ala	Val	Pro	Thr	Thr	Leu		50
	20				25							30					

Pro Pro Asn Leu Leu Glu Leu Tyr Leu Tyr Asn Asn Ile Ile Lys Lys  
35 40 45

Ile Gln Glu Asn Asp Phe Asn Asn Leu Asn Glu Leu Gln Val Leu Asp  
5 50 55 60

Leu Arg Gly Asn Cys Pro Arg Cys His Asn Val Pro Tyr Pro Cys Thr  
65 70 75 80

10 Pro Cys Glu Asn Asn Ser Pro Leu Gln Ile His Asp Asn Ala Phe Asn.  
85 90 95

Ser Ser Thr Asp  
15 100

## 15 (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- 20 (A) LENGTH: 1756 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
- 30 (A) NAME/KEY: CDS  
(B) LOCATION: 1..1182
- (ix) FEATURE:
- 35 (A) NAME/KEY: misc\_feature  
(B) LOCATION: 1643  
(D) OTHER INFORMATION: /note= "nucleotide 1643 designated A, may be A or G"
- (ix) FEATURE:
- 40 (A) NAME/KEY: misc\_feature  
(B) LOCATION: 1664  
(D) OTHER INFORMATION: /note= "nucleotide 1664 designated C, may be A, C, G, or T"
- (ix) FEATURE:
- 45 (A) NAME/KEY: misc\_feature  
(B) LOCATION: 1680  
(D) OTHER INFORMATION: /note= "nucleotides 1680 and 1735 designated G, may be G or T"
- (ix) FEATURE:
- 50 (A) NAME/KEY: misc\_feature  
(B) LOCATION: 1719  
(D) OTHER INFORMATION: /note= "nucleotide 1719 designated C, may be C or T"
- 55 (ix) FEATURE:
- (A) NAME/KEY: misc\_feature  
(B) LOCATION: 1727  
(D) OTHER INFORMATION: /note= "nucleotide 1727 designated A, may be A, G, or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

	TCT CCA GAA ATT CCC TGG AAT TCC TTG CCT CCT GAG GTT TTT GAG GGT	48
5	Ser Pro Glu Ile Pro Trp Asn Ser Leu Pro Pro Glu Val Phe Glu Gly 1 5 10 15	
	ATG CCG CCA AAT CTA AAG AAT CTC TCC TTG GCC AAA AAT GGG CTC AAA	96
	Met Pro Pro Asn Leu Lys Asn Leu Ser Leu Ala Lys Asn Gly Leu Lys 20 25 30	
10	TCT TTC TTT TGG GAC AGA CTC CAG TTA CTG AAG CAT TTG GAA ATT TTG	144
	Ser Phe Phe Trp Asp Arg Leu Gln Leu Leu Lys His Leu Glu Ile Leu 35 40 45	
15	GAC CTC AGC CAT AAC CAG CTG ACA AAA GTA CCT GAG AGA TTG GCC AAC	192
	Asp Leu Ser His Asn Gln Leu Thr Lys Val Pro Glu Arg Leu Ala Asn 50 55 60	
20	TGT TCC AAA AGT CTC ACA ACA CTG ATT CTT AAG CAT AAT CAA ATC AGG	240
	Cys Ser Lys Ser Leu Thr Thr Leu Ile Leu Lys His Asn Gln Ile Arg 65 70 75 80	
25	CAA TTG ACA AAA TAT TTT CTA GAA GAT GCT TTG CAA TTG CGC TAT CTA	288
	Gln Leu Thr Lys Tyr Phe Leu Glu Asp Ala Leu Gln Leu Arg Tyr Leu 85 90 95	
	GAC ATC AGT TCA AAT AAA ATC CAG GTC ATT CAG AAG ACT AGC TTC CCA	336
	Asp Ile Ser Ser Asn Lys Ile Gln Val Ile Gln Lys Thr Ser Phe Pro 100 105 110	
30	GAA AAT GTC CTC AAC AAT CTG GAG ATG TTG GTT TTA CAT CAC AAT CGC	384
	Glu Asn Val Leu Asn Asn Leu Glu Met Leu Val Leu His His Asn Arg 115 120 125	
35	TTT CTT TGC AAC TGT GAT GCT GTG TGG TTT GTC TGG TGG GTT AAC CAT	432
	Phe Leu Cys Asn Cys Asp Ala Val Trp Phe Val Trp Trp Val Asn His 130 135 140	
40	ACA GAT GTT ACT ATT CCA TAC CTG GCC ACT GAT GTG ACT TGT GTA GGT	480
	Thr Asp Val Thr Ile Pro Tyr Leu Ala Thr Asp Val Thr Cys Val Gly 145 150 155 160	
45	CCA GGA GCA CAC AAA GGT CAA AGT GTC ATA TCC CTT GAT CTG TAT ACG	528
	Pro Gly Ala His Lys Gly Gln Ser Val Ile Ser Leu Asp Leu Tyr Thr 165 170 175	
	TGT GAG TTA GAT CTC ACA AAC CTG ATT CTG TTC TCA GTT TCC ATA TCA	576
	Cys Glu Leu Asp Leu Thr Asn Leu Ile Leu Phe Ser Val Ser Ile Ser 180 185 190	
50	TCA GTC CTC TTT CTT ATG GTA GTT ATG ACA ACA AGT CAC CTC TTT TTC	624
	Ser Val Leu Phe Leu Met Val Val Met Thr Thr Ser His Leu Phe Phe 195 200 205	
55	TGG GAT ATG TGG TAC ATT TAT TAT TTT TGG AAA GCA AAG ATA AAG GGG	672
	Trp Asp Met Trp Tyr Ile Tyr Tyr Phe Trp Lys Ala Lys Ile Lys Gly 210 215 220	
60	TAT CCA GCA TCT GCA ATC CCA TGG AGT CCT TGT TAT GAT GCT TTT ATT	720
	Tyr Pro Ala Ser Ala Ile Pro Trp Ser Pro Cys Tyr Asp Ala Phe Ile 225 230 235 240	

	GTG TAT GAC ACT AAA AAC TCA GCT GTG ACA GAA TGG GTT TTG CAG GAG Val Tyr Asp Thr Lys Asn Ser Ala Val Thr Glu Trp Val Leu Gln Glu 245 250 255	768
5	CTG GTG GCA AAA TTG GAA GAT CCA AGA GAA AAA CAC TTC AAT TTG TGT Leu Val Ala Lys Leu Glu Asp Pro Arg Glu Lys His Phe Asn Leu Cys 260 265 270	816
10	CTA GAA GAA AGA GAC TGG CTA CCA GGA CAG CCA GTT CTA GAA AAC CTT. Leu Glu Glu Arg Asp Trp Leu Pro Gly Gln Pro Val Leu Glu Asn Leu 275 280 285	864
15	TCC CAG AGC ATA CAG CTC AGC AAA AAG ACA GTG TTT GTG ATG ACA CAG Ser Gln Ser Ile Gln Leu Ser Lys Lys Thr Val Phe Val Met Thr Gln 290 295 300	912
20	AAA TAT GCT AAG ACT GAG AGT TTT AAG ATG GCA TTT TAT TTG TCT CAT Lys Tyr Ala Lys Thr Glu Ser Phe Lys Met Ala Phe Tyr Leu Ser His 305 310 315 320	960
25	CAG AGG CTC CTG GAT GAA AAA GTG GAT GTG ATT ATC TTG ATA TTC TTG Gln Arg Leu Leu Asp Glu Lys Val Asp Val Ile Ile Leu Ile Phe Leu 325 330 335	1008
30	GAA AGA CCT CTT CAG AAG TCT AAG TTT CTT CAG CTC AGG AAG AGA CTC Glu Arg Pro Leu Gln Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg Leu 340 345 350	1056
35	TGC AGG AGC TCT GTC CTT GAG TGG CCT GCA AAT CCA CAG GCT CAC CCA Cys Arg Ser Ser Val Leu Glu Trp Pro Ala Asn Pro Gln Ala His Pro 355 360 365	1104
40	TAC TTC TGG CAG TGC CTG AAA AAT GCC CTG ACC ACA GAC AAT CAT GTG Tyr Phe Trp Gln Cys Leu Lys Asn Ala Leu Thr Thr Asp Asn His Val 370 375 380	1152
45	GCT TAT AGT CAA ATG TTC AAG GAA ACA GTC TAGCTCTCTG AAGAATGTCA Ala Tyr Ser Gln Met Phe Lys Glu Thr Val 385 390	1202
50	CCACCTAGGA CATGCCTTGG TACCTGAAGT TTTCATCAAAG GTTTCCATAA ATGAAGGTCT GAATTTTCC TAACAGTTGT CATGGCTCAG ATTGGTGGGA AATCATCAAT ATATGGCTAA GAAATTAAGA AGGGGAGACT GATAGAAGAT AATTCTTTC TTCATGTGCC ATGCTCAGTT AAATATTCCT CCTAGCTCAA ATCTGAAAAA CTGTGCCTAG GAGACAACAC AAGGCTTGA AGTAGTAGGG TAAAAATTAC ACAAGCTTTC TCTCTCTCTG ATACTGAAC GTACCAGAGT TCAATGAAAT AAAAGCCCAG AGAACTTCTC AGTAAATGGT TTCATTATCA TGTAGTATCC ACCATGCAAT ATGCCACAAA ACCGCTACTG GTACAGGACA GCTGGTAGCT GCTTCAAGGC CTCTTATCAT TTTCTTGGGG CCCATGGAGG GGTTCTCTGG GAAAAAGGGA AGGTTTTTT TGGCCATCCA TGAA	1262 1322 1382 1442 1502 1562 1622 1682 1742 1756

## (2) INFORMATION FOR SEQ ID NO:30:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 394 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: protein

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ser Pro Glu Ile Pro Trp Asn Ser Leu Pro Pro Glu Val Phe Glu Gly  
1 5 10 15

Met Pro Pro Asn Leu Lys Asn Leu Ser Leu Ala Lys Asn Gly Leu Lys  
20 25 30

Ser Phe Phe Trp Asp Arg Leu Gln Leu Leu Lys His Leu Glu Ile Leu  
20 35 40 45

Asp Leu Ser His Asn Gln Leu Thr Lys Val Pro Glu Arg Leu Ala Asn  
50 55 60

25 Cys Ser Lys Ser Leu Thr Thr Leu Ile Leu Lys His Asn Gln Ile Arg  
65 70 75 80

Gln Leu Thr Lys Tyr Phe Leu Glu Asp Ala Leu Gln Leu Arg Tyr Leu  
30 85 90 95

Asp Ile Ser Ser Asn Lys Ile Gln Val Ile Gln Lys Thr Ser Phe Pro  
100 105 110

35 Glu Asn Val Leu Asn Asn Leu Glu Met Leu Val Leu His His Asn Arg  
115 120 125

Phe Leu Cys Asn Cys Asp Ala Val Trp Phe Val Trp Trp Val Asn His  
130 135 140

40 Thr Asp Val Thr Ile Pro Tyr Leu Ala Thr Asp Val Thr Cys Val Gly  
145 150 155 160

Pro Gly Ala His Lys Gly Gln Ser Val Ile Ser Leu Asp Leu Tyr Thr  
45 165 170 175

Cys Glu Leu Asp Leu Thr Asn Leu Ile Leu Phe Ser Val Ser Ile Ser  
180 185 190

Ser Val Leu Phe Leu Met Val Val Met Thr Thr Ser His Leu Phe Phe  
50 195 200 205

Trp Asp Met Trp Tyr Ile Tyr Tyr Phe Trp Lys Ala Lys Ile Lys Gly  
210 215 220

55 Tyr Pro Ala Ser Ala Ile Pro Trp Ser Pro Cys Tyr Asp Ala Phe Ile  
225 230 235 240

Val Tyr Asp Thr Lys Asn Ser Ala Val Thr Glu Trp Val Leu Gln Glu  
60 245 250 255

Leu Val Ala Lys Leu Glu Asp Pro Arg Glu Lys His Phe Asn Leu Cys

260                  265                  270

Leu Glu Glu Arg Asp Trp Leu Pro Gly Gln Pro Val Leu Glu Asn Leu  
275                  280                  285

5                    Ser Gln Ser Ile Gln Leu Ser Lys Lys Thr Val Phe Val Met Thr Gln  
                      290                295                300

10                  Lys Tyr Ala Lys Thr Glu Ser Phe Lys Met Ala Phe Tyr Leu Ser His  
                      305                310                315                320

                      Gln Arg Leu Leu Asp Glu Lys Val Asp Val Ile Ile Leu Ile Phe Leu  
                      325                330                335

15                  Glu Arg Pro Leu Gln Lys Ser Lys Phe Leu Gln Leu Arg Lys Arg Leu  
                      340                345                350

                      Cys Arg Ser Ser Val Leu Glu Trp Pro Ala Asn Pro Gln Ala His Pro  
                      355                360                365

20                  Tyr Phe Trp Gln Cys Leu Lys Asn Ala Leu Thr Thr Asp Asn His Val  
                      370                375                380

                      Ala Tyr Ser Gln Met Phe Lys Glu Thr Val  
                      385                390

## (2) INFORMATION FOR SEQ ID NO:31:

- 30                  (i) SEQUENCE CHARACTERISTICS:  
                      (A) LENGTH: 999 base pairs  
                      (B) TYPE: nucleic acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear
- 35                  (ii) MOLECULE TYPE: cDNA
- 40                  (ix) FEATURE:  
                      (A) NAME/KEY: CDS  
                      (B) LOCATION: 2..847
- 45                  (ix) FEATURE:  
                      (A) NAME/KEY: misc\_feature  
                      (B) LOCATION: 4  
                      (D) OTHER INFORMATION: /note= "nucleotides 4 and 23  
                      designated C, each may be A, C, G, or T"
- 50                  (ix) FEATURE:  
                      (A) NAME/KEY: misc\_feature  
                      (B) LOCATION: 650  
                      (D) OTHER INFORMATION: /note= "nucleotide 650 designated  
                      G, may be A or G"
- 55                  (ix) FEATURE:  
                      (A) NAME/KEY: misc\_feature  
                      (B) LOCATION: 715  
                      (D) OTHER INFORMATION: /note= "nucleotides 715, 825, and  
                      845 designated C, each may be C or T"
- 60                  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

	C TCC GAT GCC AAG ATT CGG CAC CAG GCA TAT TCA GAG GTC ATG ATG Ser Asp Ala Lys Ile Arg His Gln Ala Tyr Ser Glu Val Met Met 1 5 10 15	46
5	GTT GGA TGG TCA GAT TCA TAC ACC TGT GAA TAC CCT TTA AAC CTA AGG Val Gly Trp Ser Asp Ser Tyr Thr Cys Glu Tyr Pro Leu Asn Leu Arg 20 25 30	94
10	GGA ACT AGG TTA AAA GAC GTT CAT CTC CAC GAA TTA TCT TGC AAC ACA Gly Thr Arg Leu Lys Asp Val His Leu His Glu Leu Ser Cys Asn Thr 35 40 45	142
15	GCT CTG TTG ATT GTC ACC ATT GTG GTT ATT ATG CTA GTT CTG GGG TTG Ala Leu Leu Ile Val Thr Ile Val Val Ile Met Leu Val Leu Gly Leu 50 55 60	190
20	GCT GTG GCC TTC TGC TGT CTC CAC TTT GAT CTG CCC TGG TAT CTC AGG Ala Val Ala Phe Cys Cys Leu His Phe Asp Leu Pro Trp Tyr Leu Arg 65 70 75	238
25	ATG CTA GGT CAA TGC ACA CAA ACA TGG CAC AGG GTT AGG AAA ACA ACC Met Leu Gly Gln Cys Thr Gln Thr Trp His Arg Val Arg Lys Thr Thr 80 85 90 95	286
30	CAA GAA CAA CTC AAG AGA AAT GTC CGA TTC CAC GCA TTT ATT TCA TAC Gln Glu Gln Leu Lys Arg Asn Val Arg Phe His Ala Phe Ile Ser Tyr 100 105 110	334
35	AGT GAA CAT GAT TCT CTG TGG GTG AAG AAT GAA TTG ATC CCC AAT CTA Ser Glu His Asp Ser Leu Trp Val Lys Asn Glu Leu Ile Pro Asn Leu 115 120 125	382
40	GAG AAG GAA GAT GGT TCT ATC TTG ATT TGC CTT TAT GAA AGC TAC TTT Glu Lys Glu Asp Gly Ser Ile Leu Ile Cys Leu Tyr Glu Ser Tyr Phe 130 135 140	430
45	GAC CCT GGC AAA AGC ATT AGT GAA AAT ATT GTA AGC TTC ATT GAG AAA Asp Pro Gly Lys Ser Ile Ser Glu Asn Ile Val Ser Phe Ile Glu Lys 145 150 155	478
50	AGC TAT AAG TCC ATC TTT GTT TTG TCT CCC AAC TTT GTC CAG AAT GAG Ser Tyr Lys Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Asn Glu 160 165 170 175	526
55	TGG TGC CAT TAT GAA TTC TAC TTT GCC CAC CAC AAT CTC TTC CAT GAA Trp Cys His Tyr Glu Phe Tyr Phe Ala His His Asn Leu Phe His Glu 180 185 190	574
60	AAT TCT GAT CAC ATA ATT CTT ATC TTA CTG GAA CCC ATT CCA TTC TAT Asn Ser Asp His Ile Ile Leu Ile Leu Glu Pro Ile Pro Phe Tyr 195 200 205	622
	TGC ATT CCC ACC AGG TAT CAT AAA CTG GAA GCT CTC CTG GAA AAA AAA Cys Ile Pro Thr Arg Tyr His Lys Leu Glu Ala Leu Leu Glu Lys Lys 210 215 220	670
	GCA TAC TTG GAA TGG CCC AAG GAT AGG CGT AAA TGT GGG CTT TTC TGG Ala Tyr Leu Glu Trp Pro Lys Asp Arg Arg Lys Cys Gly Leu Phe Trp 225 230 235	718

	GCA AAC CTT CGA GCT GCT GTT AAT GTT AAT GTA TTA GCC ACC AGA GAA Ala Asn Leu Arg Ala Ala Val Asn Val Asn Val Leu Ala Thr Arg Glu 240                   245                   250                   255	766
5	ATG TAT GAA CTG CAG ACA TTC ACA GAG TTA AAT GAA GAG TCT CGA GGT Met Tyr Glu Leu Gln Thr Phe Thr Glu Leu Asn Glu Glu Ser Arg Gly 260                   265                   270	814
10	TCT ACA ATC TCT CTG ATG AGA ACA GAC TGT CTA TAAAATCCCA CAGTCCTTGG Ser Thr Ile Ser Leu Met Arg Thr Asp Cys Leu 275                   280	867
	GAAGTTGGGG ACCACATACA CTGTTGGAT GTACATTGAT ACAACCTTA TGATGCCAAT	927
15	TTGACAATAT TTATTAAAAT AAAAAATGGT TATTCCCTTC AAAAAAAA AAAAAAAA AAAAAAA AA	987
		999

## 20 (2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

30	Ser Asp Ala Lys Ile Arg His Gln Ala Tyr Ser Glu Val Met Met Val 1                   5                   10                   15
35	Gly Trp Ser Asp Ser Tyr Thr Cys Glu Tyr Pro Leu Asn Leu Arg Gly 20                   25                   30
	Thr Arg Leu Lys Asp Val His Leu His Glu Leu Ser Cys Asn Thr Ala 35                   40                   45
40	Leu Leu Ile Val Thr Ile Val Val Ile Met Leu Val Leu Gly Leu Ala 50                   55                   60
	Val Ala Phe Cys Cys Leu His Phe Asp Leu Pro Trp Tyr Leu Arg Met 65                   70                   75                   80
45	Leu Gly Gln Cys Thr Gln Thr Trp His Arg Val Arg Lys Thr Thr Gln 85                   90                   95
50	Glu Gln Leu Lys Arg Asn Val Arg Phe His Ala Phe Ile Ser Tyr Ser 100               105               110
	Glu His Asp Ser Leu Trp Val Lys Asn Glu Leu Ile Pro Asn Leu Glu 115               120               125
55	Lys Glu Asp Gly Ser Ile Leu Ile Cys Leu Tyr Glu Ser Tyr Phe Asp 130               135               140
	Pro Gly Lys Ser Ile Ser Glu Asn Ile Val Ser Phe Ile Glu Lys Ser 145               150               155               160
60	Tyr Lys Ser Ile Phe Val Leu Ser Pro Asn Phe Val Gln Asn Glu Trp

	165	170	175
	Cys His Tyr Glu Phe Tyr Phe Ala His His Asn Leu Phe His Glu Asn		
	180	185	190
5	Ser Asp His Ile Ile Leu Ile Leu Leu Glu Pro Ile Pro Phe Tyr Cys		
	195	200	205
10	Ile Pro Thr Arg Tyr His Lys Leu Glu Ala Leu Leu Glu Lys Lys Ala		
	210	215	220
	Tyr Leu Glu Trp Pro Lys Asp Arg Arg Lys Cys Gly Leu Phe Trp Ala		
	225	230	235
			240
15	Asn Leu Arg Ala Ala Val Asn Val Asn Val Leu Ala Thr Arg Glu Met		
	245	250	255
	Tyr Glu Leu Gln Thr Phe Thr Glu Leu Asn Glu Glu Ser Arg Gly Ser		
20	260	265	270
	Thr Ile Ser Leu Met Arg Thr Asp Cys Leu		
	275	280	

(2) INFORMATION FOR SEQ ID NO:33:

25	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1173 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
30	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: cDNA
35	(ix) FEATURE:
	(A) NAME/KEY: CDS
	(B) LOCATION: 1..1008
40	(ix) FEATURE:
	(A) NAME/KEY: misc_feature
	(B) LOCATION: 854
	(D) OTHER INFORMATION: /note= "nucleotide 854 designated A, may be A or T"
45	(ix) FEATURE:
	(A) NAME/KEY: misc_feature
	(B) LOCATION: 1171
	(D) OTHER INFORMATION: /note= "nucleotides 1171 and 1172 designated C, each may be A, C, G, or T"
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

55	CTG CCT GCT GGC ACC CGG CTC CGG AGG CTG GAT GTC AGC TGC AAC AGC Leu Pro Ala Gly Thr Arg Leu Arg Arg Leu Asp Val Ser Cys Asn Ser	48
	1               5               10              15	
60	ATC AGC TTC GTG GCC CCC GGC TTC TTT TCC AAG GCC AAG GAG CTG CGA Ile Ser Phe Val Ala Pro Gly Phe Phe Ser Lys Ala Lys Glu Leu Arg	96
	20              25              30	

	GAG CTC AAC CTT AGC GCC AAC GCC CTC AAG ACA GTG GAC CAC TCC TGG Glu Leu Asn Leu Ser Ala Asn Ala Leu Lys Thr Val Asp His Ser Trp 35 40 45	144
5	TTT GGG CCC CTG GCG AGT GCC CTG CAA ATA CTA GAT GTA AGC GCC AAC Phe Gly Pro Leu Ala Ser Ala Leu Gln Ile Leu Asp Val Ser Ala Asn 50 55 60	192
10	CCT CTG CAC TGC GCC TGT GGG GCG GCC TTT ATG GAC TTC CTG CTG GAG Pro Leu His Cys Ala Cys Gly Ala Ala Phe Met Asp Phe Leu Leu Glu 65 70 75 80	240
15	GTG CAG GCT GCC GTG CCC GGT CTG CCC AGC CGG GTG AAG TGT GGC AGT Val Gln Ala Ala Val Pro Gly Leu Pro Ser Arg Val Lys Cys Gly Ser 85 90 95	288
20	CCG GGC CAG CTC CAG GGC CTC AGC ATC TTT GCA CAG GAC CTG CGC CTC Pro Gly Gln Leu Gln Gly Leu Ser Ile Phe Ala Gln Asp Leu Arg Leu 100 105 110	336
25	TGC CTG GAT GAG GCC CTC TCC TGG GAC TGT TTC GCC CTC TCG CTG CTG Cys Leu Asp Glu Ala Leu Ser Trp Asp Cys Phe Ala Leu Ser Leu Leu 115 120 125	384
30	GCT GTG GCT CTG GGC CTG GGT GTG CCC ATG CTG CAT CAC CTC TGT GGC Ala Val Ala Leu Gly Leu Gly Val Pro Met Leu His His Leu Cys Gly 130 135 140	432
35	TGG GAC CTC TGG TAC TGC TTC CAC CTG TGC CTG GCC TGG CTT CCC TGG Trp Asp Leu Trp Tyr Cys Phe His Leu Cys Leu Ala Trp Leu Pro Trp 145 150 155 160	480
40	CGG GGG CGG CAA AGT GGG CGA GAT GAG GAT GCC CTG CCC TAC GAT GCC Arg Gly Arg Gln Ser Gly Arg Asp Glu Asp Ala Leu Pro Tyr Asp Ala 165 170 175	528
45	TTC GTG GTC TTC GAC AAA ACG CAG AGC GCA GTG GCA GAC TGG GTG TAC Phe Val Val Phe Asp Lys Thr Gln Ser Ala Val Ala Asp Trp Val Tyr 180 185 190	576
50	AAC GAG CTT CGG GGG CAG CTG GAG GAG TGC CGT GGG CGC TGG GCA CTC Asn Glu Leu Arg Gly Gln Leu Glu Glu Cys Arg Gly Arg Trp Ala Leu 195 200 205	624
55	CGC CTG TGC CTG GAG GAA CGC GAC TGG CTG CCT GGC AAA ACC CTC TTT Arg Leu Cys Leu Glu Glu Arg Asp Trp Leu Pro Gly Lys Thr Leu Phe 210 215 220	672
60	GAG AAC CTG TGG GCC TCG GTC TAT GGC AGC CGC AAG ACG CTG TTT GTG Glu Asn Leu Trp Ala Ser Val Tyr Gly Ser Arg Lys Thr Leu Phe Val 225 230 235 240	720
	CTG GCC CAC ACG GAC CGG GTC AGT GGT CTC TTG CGC GCC AGC TTC CTG Leu Ala His Thr Asp Arg Val Ser Gly Leu Leu Arg Ala Ser Phe Leu 245 250 255	768
	CTG GCC CAG CAG CGC CTG CTG GAG GAC CGC AAG GAC GTC GTG GTG CTG Leu Ala Gln Gln Arg Leu Leu Glu Asp Arg Lys Asp Val Val Val Leu 260 265 270	816
	GTG ATC CTG AGC CCT GAC GGC CGC CGC TCC CGC TAC GAG CGG CTG CGC	864

	Val Ile Leu Ser Pro Asp Gly Arg Arg Ser Arg Tyr Glu Arg Leu Arg			
	275	280	285	
5	CAG CGC CTC TGC CGC CAG AGT GTC CTC CTC TGG CCC CAC CAG CCC AGT Gln Arg Leu Cys Arg Gln Ser Val Leu Leu Trp Pro His Gln Pro Ser		912	
	290	295	300	
10	GGT CAG CGC AGC TTC TGG GCC CAG CTG GGC ATG GCC CTG ACC AGG GAC Gly Gln Arg Ser Phe Trp Ala Gln Leu Gly Met Ala Leu Thr Arg Asp		960	
	305	310	315	320
15	AAC CAC CAC TTC TAT AAC CGG AAC TTC TGC CAG GGA CCC ACG GCC GAA Asn His His Phe Tyr Asn Arg Asn Phe Cys Gln Gly Pro Thr Ala Glu		1008	
	325	330	335	
20	TAGCCGTGAG CCGGAATCCT GCACGGTGCC ACCTCCACAC TCACCTCACCC TCTGCCTGCC TGGTCTGACC CTCCCCCTGCT CGCCTCCCTC ACCCCACACC TGACACAGAG CAGGCACTCA		1068	
	ATAAATGCTA CCGAAGGCTA AAAAAAAAAA AAAAAAAAAA AACCA		1128	
			1173	

## (2) INFORMATION FOR SEQ ID NO:34:

25	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 336 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
30	(ii) MOLECULE TYPE: protein		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:		
35	Leu Pro Ala Gly Thr Arg Leu Arg Arg Leu Asp Val Ser Cys Asn Ser		
	1 5 10 15		
	Ile Ser Phe Val Ala Pro Gly Phe Phe Ser Lys Ala Lys Glu Leu Arg		
	20 25 30		
40	Glu Leu Asn Leu Ser Ala Asn Ala Leu Lys Thr Val Asp His Ser Trp		
	35 40 45		
	Phe Gly Pro Leu Ala Ser Ala Leu Gln Ile Leu Asp Val Ser Ala Asn		
	50 55 60		
45	Pro Leu His Cys Ala Cys Gly Ala Ala Phe Met Asp Phe Leu Leu Glu		
	65 70 75 80		
50	Val Gln Ala Ala Val Pro Gly Leu Pro Ser Arg Val Lys Cys Gly Ser		
	85 90 95		
	Pro Gly Gln Leu Gln Gly Leu Ser Ile Phe Ala Gln Asp Leu Arg Leu		
	100 105 110		
55	Cys Leu Asp Glu Ala Leu Ser Trp Asp Cys Phe Ala Leu Ser Leu Leu		
	115 120 125		
	Ala Val Ala Leu Gly Leu Gly Val Pro Met Leu His His Leu Cys Gly		
	130 135 140		
60	Trp Asp Leu Trp Tyr Cys Phe His Leu Cys Leu Ala Trp Leu Pro Trp		

	145	150	155	160
	Arg Gly Arg Gln Ser Gly Arg Asp Glu Asp Ala Leu Pro Tyr Asp Ala			
	165	170	175	
5	Phe Val Val Phe Asp Lys Thr Gln Ser Ala Val Ala Asp Trp Val Tyr			
	180	185	190	
	Asn Glu Leu Arg Gly Gln Leu Glu Glu Cys Arg Gly Arg Trp Ala Leu			
10	195	200	205	
	Arg Leu Cys Leu Glu Glu Arg Asp Trp Leu Pro Gly Lys Thr Leu Phe			
	210	215	220	
15	Glu Asn Leu Trp Ala Ser Val Tyr Gly Ser Arg Lys Thr Leu Phe Val			
	225	230	235	240
	Leu Ala His Thr Asp Arg Val Ser Gly Leu Leu Arg Ala Ser Phe Leu			
20	245	250	255	
	Leu Ala Gln Gln Arg Leu Leu Glu Asp Arg Lys Asp Val Val Val Leu			
	260	265	270	
25	Val Ile Leu Ser Pro Asp Gly Arg Arg Ser Arg Tyr Glu Arg Leu Arg			
	275	280	285	
	Gln Arg Leu Cys Arg Gln Ser Val Leu Leu Trp Pro His Gln Pro Ser			
	290	295	300	
30	Gly Gln Arg Ser Phe Trp Ala Gln Leu Gly Met Ala Leu Thr Arg Asp			
	305	310	315	320
	Asn His His Phe Tyr Asn Arg Asn Phe Cys Gln Gly Pro Thr Ala Glu			
35	325	330	335	

## (2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

40	(A) LENGTH: 497 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

## 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGGCCCACAC	GGACCGCGTC	AGTGGCCTCC	TGCGCACCAAG	CTTCCTGCTG	GCTCAGCAGC	60	
55	GCCTGTTGGA	AGACCGCAAG	GACGTGGTGG	TGTTGGTGAT	CCTGCGTCCG	GATGCCAAC	120
	CGTCCCCGCTA	TGTGCGACTG	CGCCAGCGTC	TCTGCCGCCA	GAGTGTGCTC	TTCTGGCCCC	180
	AGCGACCCAA	CGGGCAGGGG	GGCTTCTGGG	CCCAGCTGAG	TACAGCCCTG	ACTAGGGACA	240
60	ACCGCCACTT	CTATAACCAG	AACTTCTGCC	GGGGACCTAC	AGCAGAATAG	CTCAGAGCAA	300

CAGCTGGAAA CAGCTGCATC TTCATGTCTG GTTCCCGAGT TGCTCTGCCT GCCTTGCTCT 360  
GTCTTACTAC ACCGCTATTT GGCAAGTGCG CAATATATGC TACCAAGCCA CCAGGCCAC 420  
5 GGAGCAAAGG TTGGCTGTAA AGGGTAGTTT TCTTCCCATG CATCTTCAG GAGAGTGAAG 480  
ATAGACACCA AACCCAC 497

## WHAT IS CLAIMED IS:

1. A substantially pure or recombinant DTLR2 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 4.
2. A substantially pure or recombinant DTLR3 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 6.
3. A substantially pure or recombinant DTLR4 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 26.
- 4 . A substantially pure or recombinant DTLR5 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 10.
5. A substantially pure or recombinant DTLR6 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 12.
6. A substantially pure or recombinant DTLR7 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 16 or 18.
7. A substantially pure or recombinant DTLR8 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 32.

8. A substantially pure or recombinant DTLR9 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 22.  
5
9. A substantially pure or recombinant DTLR10 protein or peptide which exhibits at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 34.  
10
10. A fusion protein comprising the protein or peptide of any of claims 1-9.
11. A binding compound which specifically binds to the protein or peptide of any of claims 1-9.  
15
12. The binding compound of claim 11 which is an antibody or antibody fragment.  
20
13. A nucleic acid encoding the protein or peptide of any of claims 1-9.
14. An expression vector comprising the nucleic acid of claim 13.  
25
15. A host cell comprising the vector of claim 14.
16. A process for recombinantly producing a polypeptide comprising culturing the host cell of claim 15 under  
30 conditions in which the polypeptide is expressed.

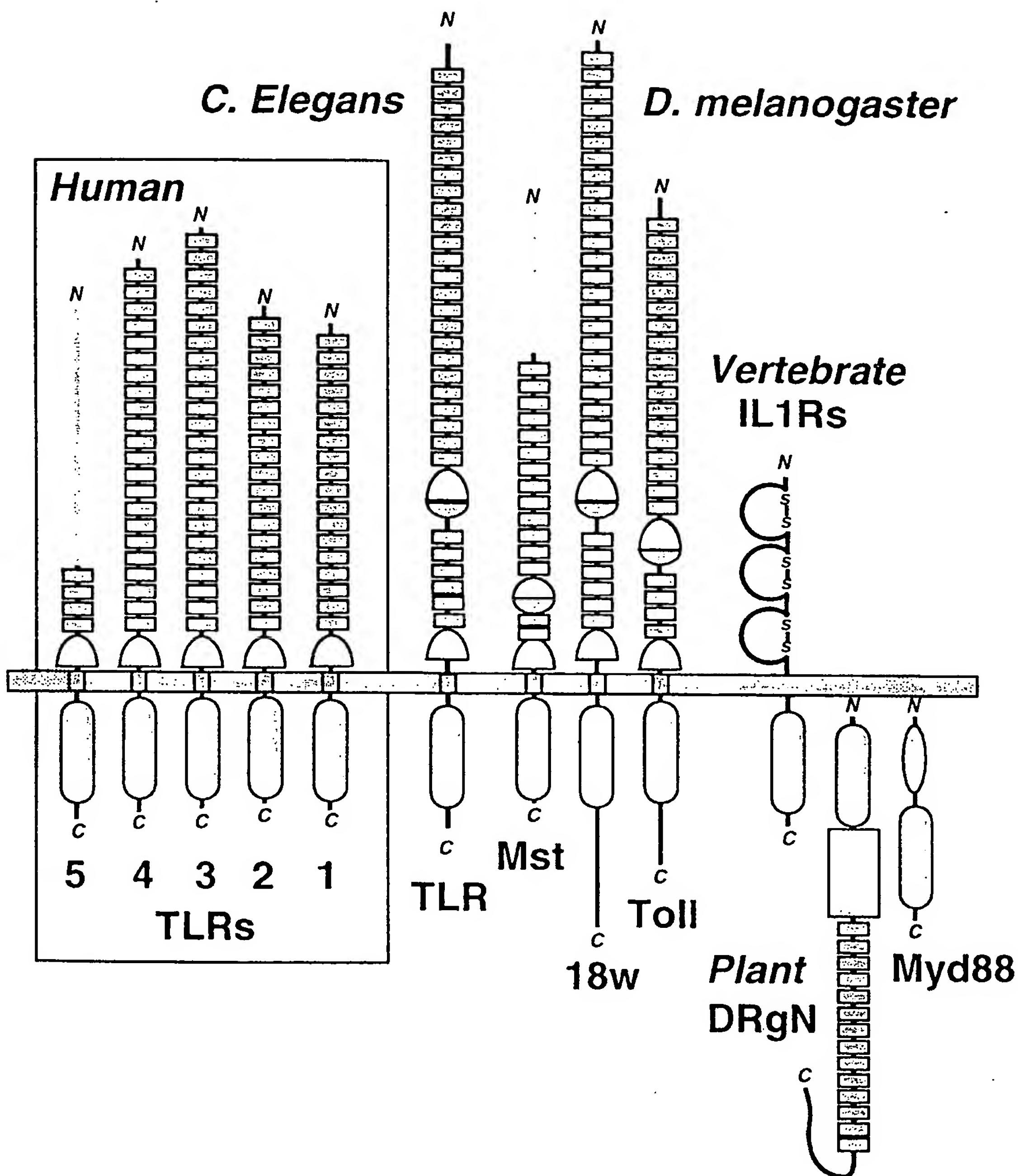
**FIG. 1**

FIG. 2A

**Block 1**

**Block 2**

**Block 3**

**Block 4**

**Block 5**

**Block 6**

**Block 7**

**Block 8**

**Block 9**

**Block 10**

ALL Cons

IL-1Rs

TLRs

PHD Sec SR

DSC Sec SR

HuL-1RS

HuL-1R3

HuL-1R4

HuL-1R6

HuL-1R7

CNL-1R1

HuMyD88

NoMyD88

HuTLR1

HuTLR2

HuTLR3

HuTLR4

HuTLR5

Dm18wh

DmMst

DmToll

ToDRgn

LULS

462 RYELESGLHEALVER..KIIILIEFTPVTDFTFPL...QSLKLLKSHRV..LKWKAQKSLSYNS...RFWKNLLYLMPAKTVKPGRDE 533

480 LLELKAGLENMASRGNHINVILVOYKAVKQMKVK....ELKRAKTVLTV..IKWKGEKSKYPO...GRFWKOLOVAMPVKKSPRWSSN 559

464 AYEQEIALHSALIONNSKVILIEMEPLGEASRLQVGDLQDSLOHLVXIOQT..IKWREDHVADKOSLSSKFWKHVAYOMPVPERASKTAS 533

465 LSEEQIAVYBALIQQGMKVILIELEKIEDQYTVM...ESIOYIKOKHGA..IRWHGQFTEQSQCMKTKFWKTVAYHMPARRCRPFLAS 549

487 SSEEQIAMYNALVODGIKVVLLELEKIEDQYEKMP...ESIKFIKOKHGA..IRWSGQFTQGPOSAKTRFWKNVRYHMPVQRSPSSKH 581

469 ASEKHLAVYNALIQQGIKILIELEKIEDQYANMP...ESIKYVKOKYGA..IRWTGQFSERSHSASTRFWKKVRYHMPSRKHQSSSGF 533

223 DFQTKFALSLSPGAHQ..KLIPIKYKAMKK...EFPSILRFITV..CDYTNPCTKS.....WFWTRLAKALSLP..... 296

181 DFQTKFALSLSPGVQO..KLIPIKYKAMKK...DFPSILRFITI..CDYTNPCTKS.....WFWTRLAKALSLP..... 243

HuTLR1

HuTLR2

HuTLR3

HuTLR4

HuTLR5

Dm18wh

DmMst

DmToll

ToDRgn

LULS

708 HYELYFAHHNLFHEQS..SLILILLEPIPOYSIPSSYH..KLKSLMARRTY..LEWPKEKSKR....GLFWANLRAAINIKLTEQAKK. 736

714 KYELDFSHFRLFDEHN..AAILILLEPIEKAIPORFC..KLAKIMNTKY..LEWPMDAEQR....EGFWVNLRAAIKS..... 734

829 RFKYHHAHQQAIEQNL..SILILYFLEEIPQYKLNHA...LCLRRGMFKSHCILNWPVQKERI....CAFAMKLOVALGIOKLCYI... 604

708 IFEYEIAOTWQFLSSRAGIIFIVLOKVENTLLRQOV..ELYRLLSRNTYL..LEWEDSVLGR....HIFWRRLRKALLDGKSWNPEGT 780

215 AFSYAQGRCLSLDLS..ALIMVVVGSLSQYOLMH...QOSIRGFYOKOQY..LRWPEDQLODV....GWFLHKSQQIKKEKEKKKDNNI 368

1117 RIEFRNAFHESLRGUA..KLVIIETTSVSAEAEDVA..ELSPYLKSVP...NRLLTQD.....HYFWEKLYRAIPIELS PRGNNY 1103

644 RLEFRALAHATSRDRC..RLIVVLYPNVKNFDSLDS...ELATYMAFNTY..LERSHP.....NFWNKLIYSMPHTKLR..... 711

832 RLEFRAAHRSALNEGR..R1IVI1VSIDIGVEKLD...ELKAYLKMNTY..LKWDGP.....WFWDKLRFALPHRRPVGNIGN 1006

84 LNELVKIMECKTRFK..QTVIPIFYDVPDSHVRNQK..ESFAKAFEEHE..TKYKDQVEG....QORWRIALNEAAHLKGSCDNAD 161

133 LMELAEIVRROEEDP..R1ILPIFYMVDPSDVRHQT...GCYKKAFRKHA..INKFDGQT.....IONWKDALKKVGDALKQWHIQKN 200

ALL Cons

IL-1Rs

TLRs

PHD Sec SR

DSC Sec SR

thEhphuhptt.bpp..DIII.bpttpp.th.....c.hphht.pph..Icaissh.t.....MFWeptithhe.

h.E.chuleashLippt..+IIIIPhtlpcn.bh.....SI+hphphhh..1+W+u-hopt.p...+FWKpIPTYIMP+

+hEhcAhpptihpppt..IIIIIPhs1phth.p...cLphhtpho..Lewps.....MFWe+pLphai..

HHHHHHHHHHH....L.EEEEEEE..LLL..LLL..HHHHHHHHH....EE..LLL..LLL..HHHHHHHHHHH..L

LHHHHHHHHHHHHHLL..LEEEEEEELLLLL..HHHHHHHHHHHHH..HEEEELLLLL..HHHHHHHHHHHLLL

BA

α1

BB

α2

BC

α3

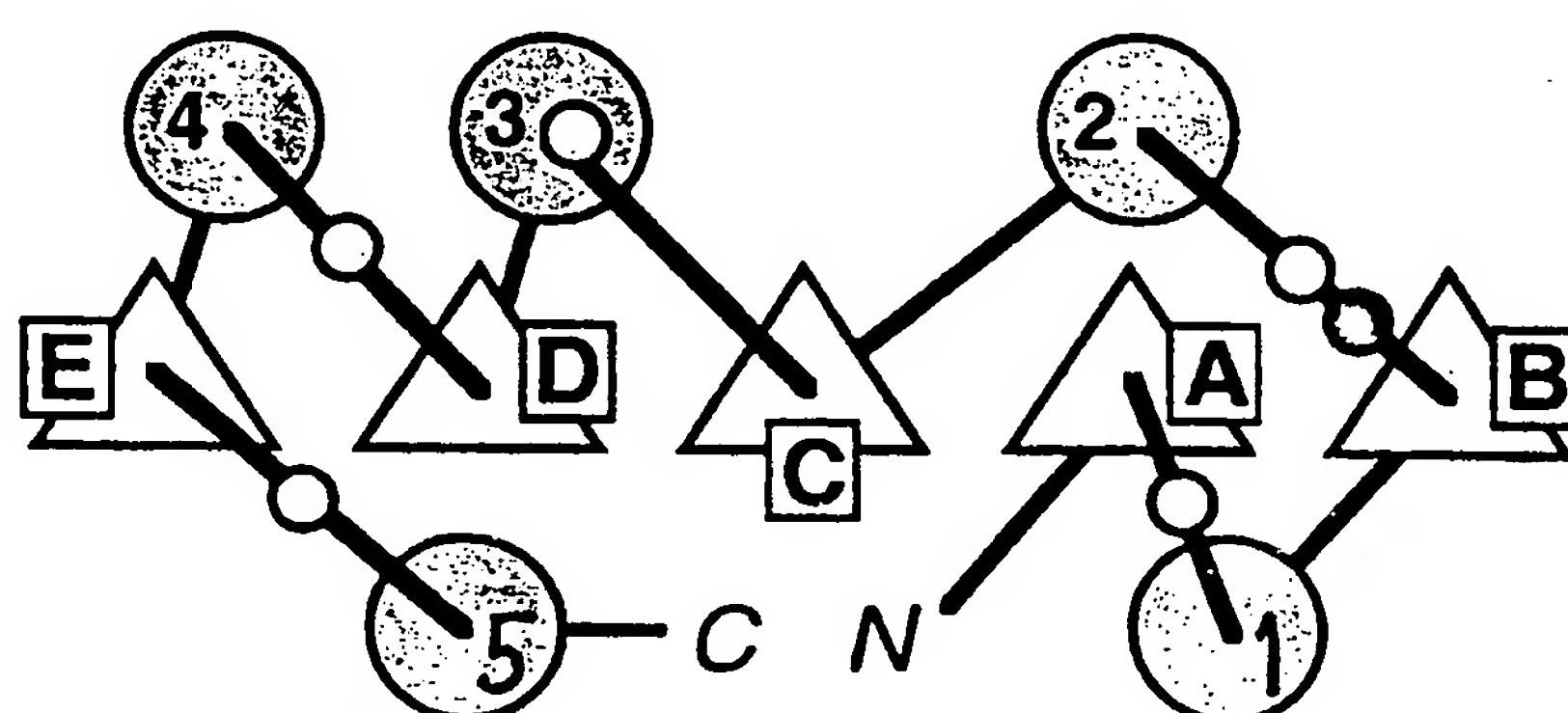
BD

α4

BE

α5

FIG. 2B



**SUBSTITUTE SHEET (RULE 26)**

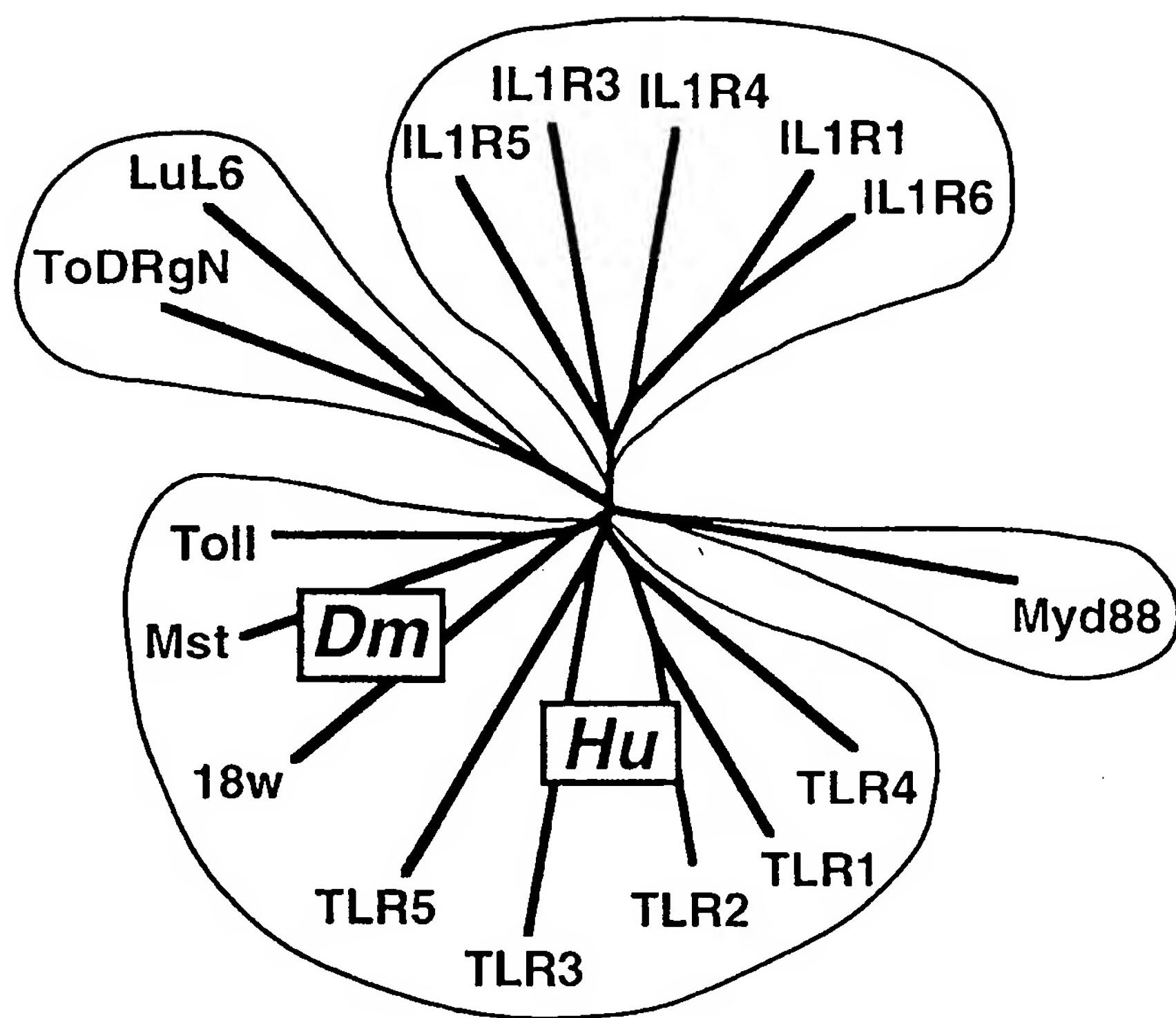


FIG. 3

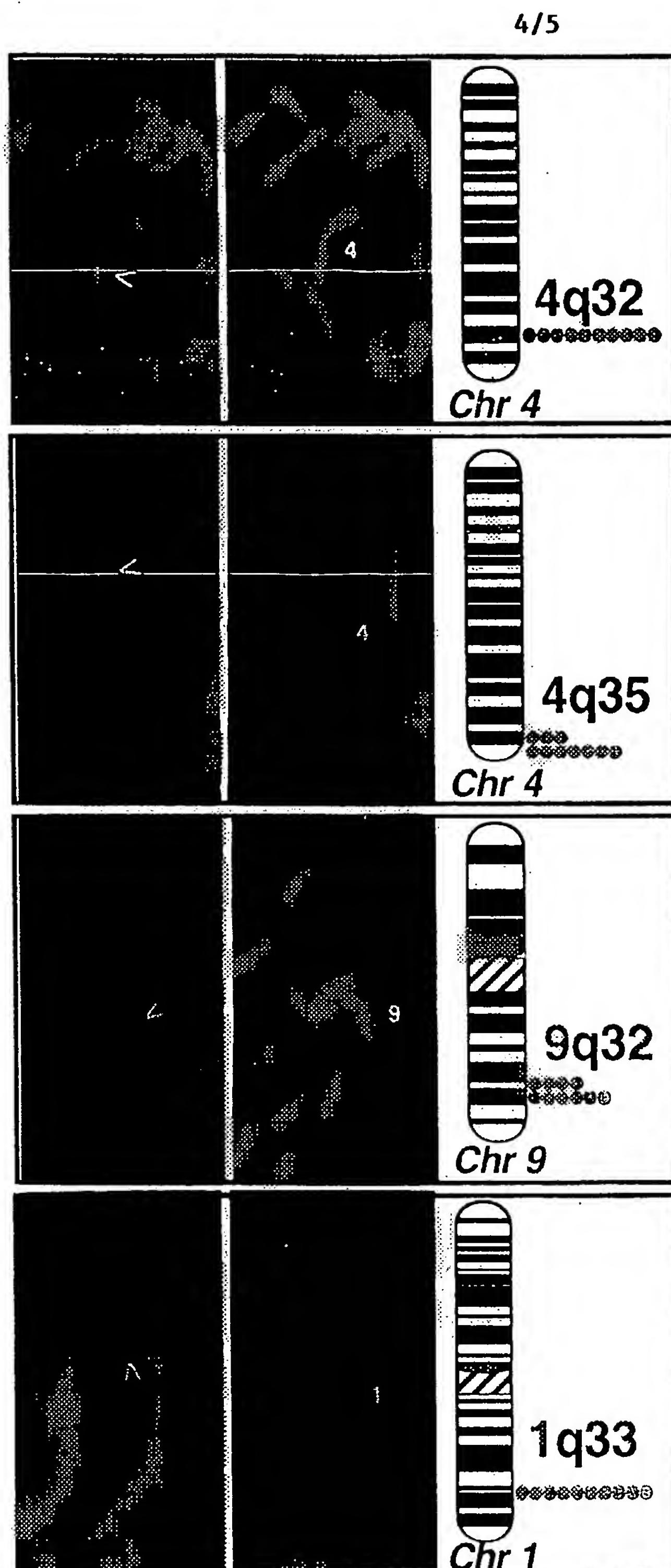


FIG. 4A

FIG. 4B

FIG. 4C

FIG. 4D

5/5

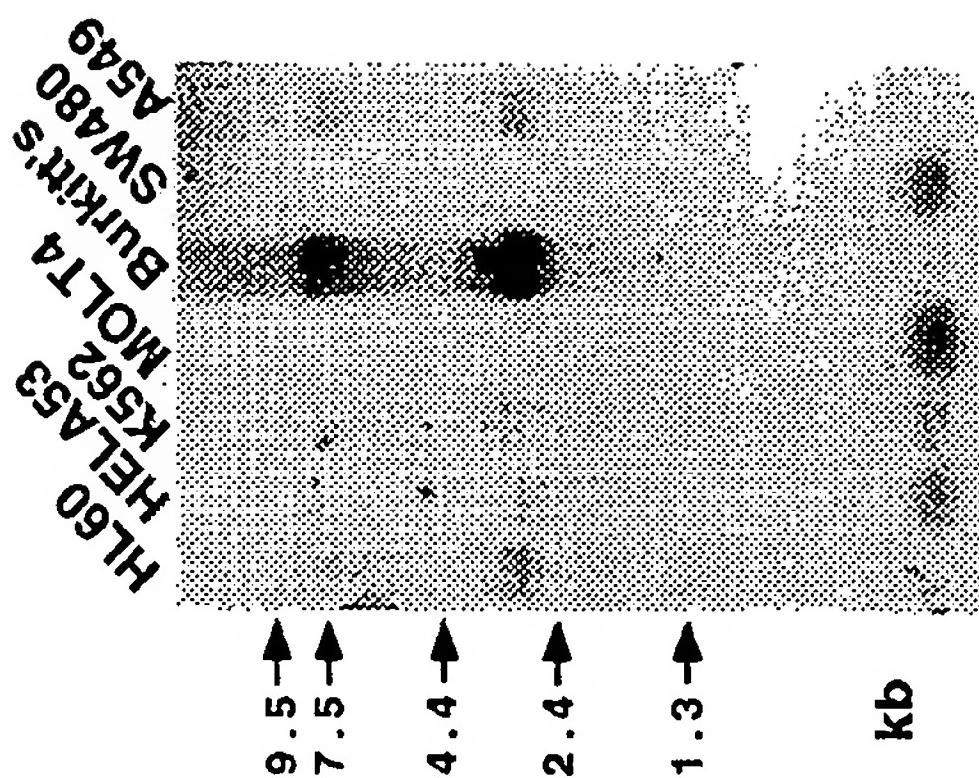


FIG. 5C

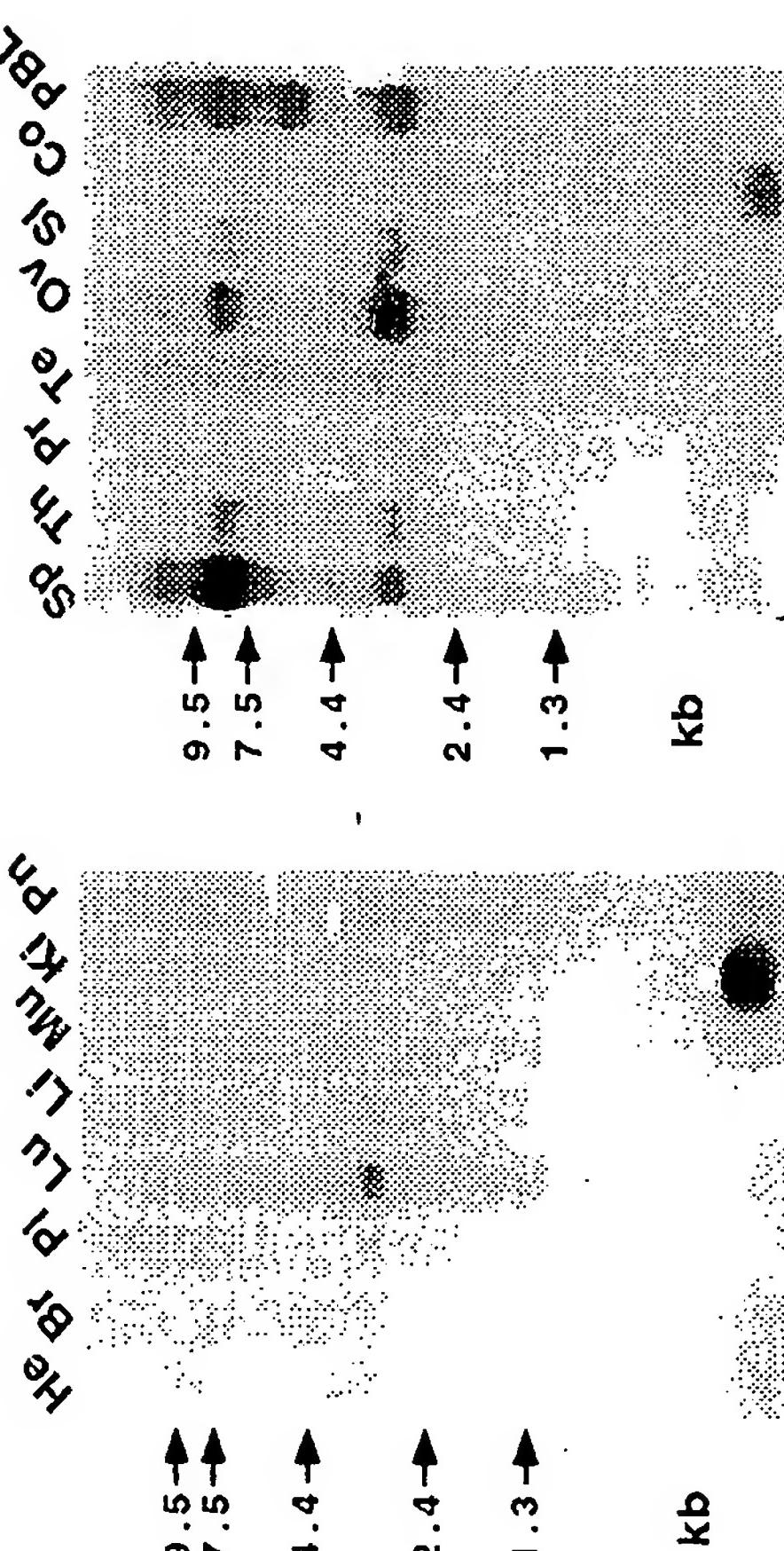


FIG. 5B

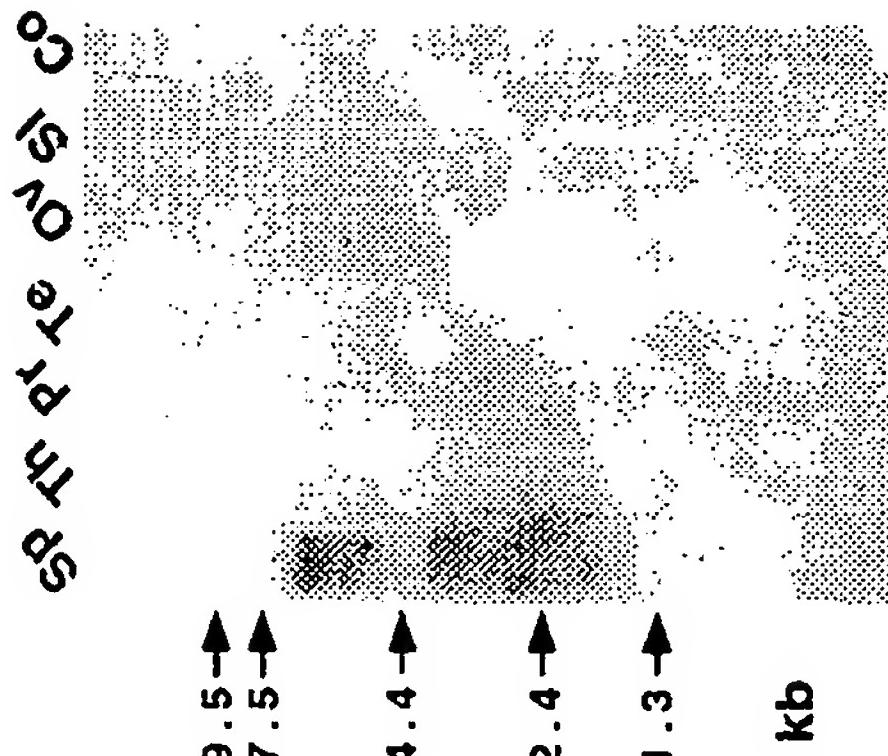


FIG. 5A

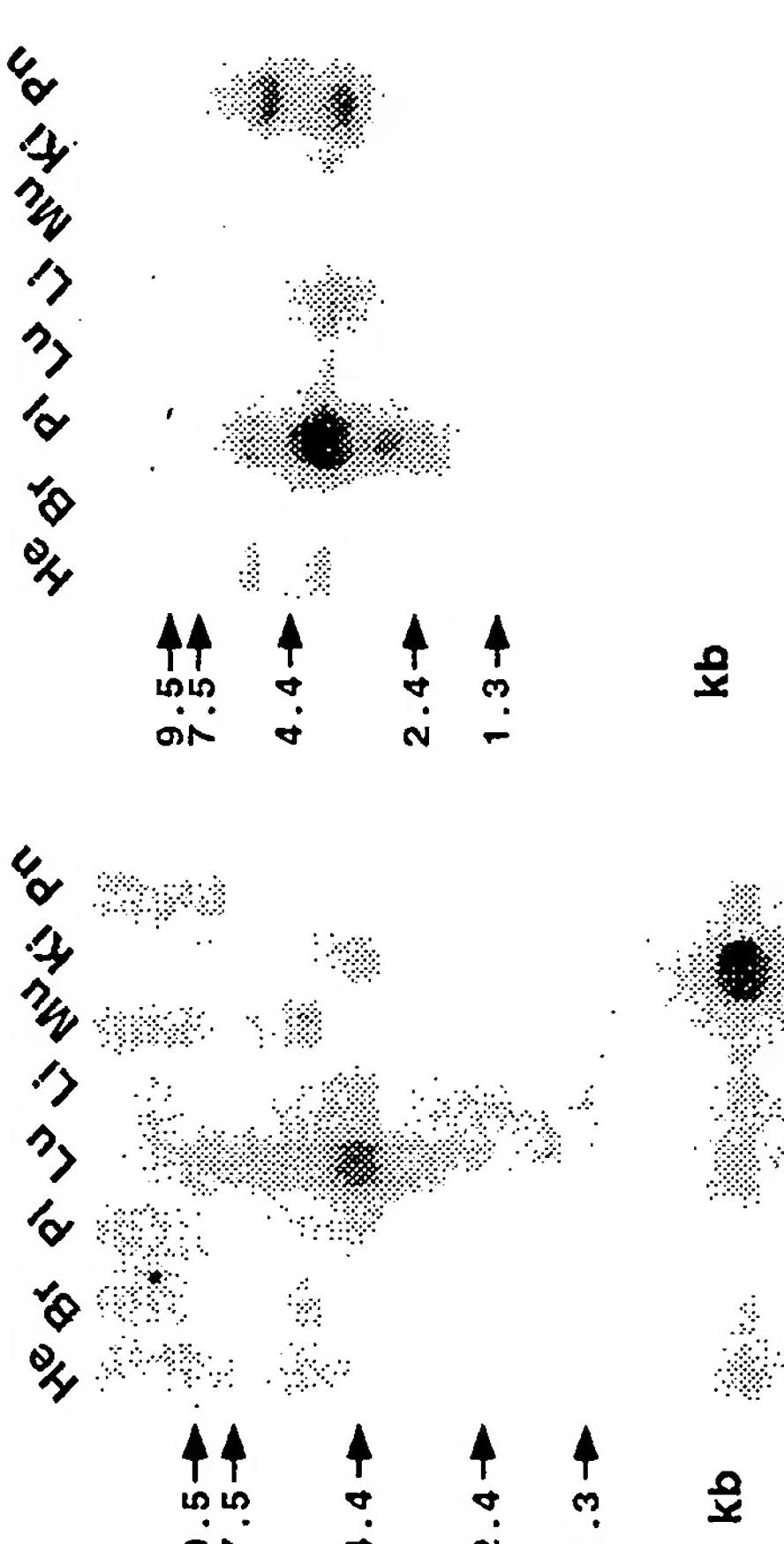


FIG. 5D

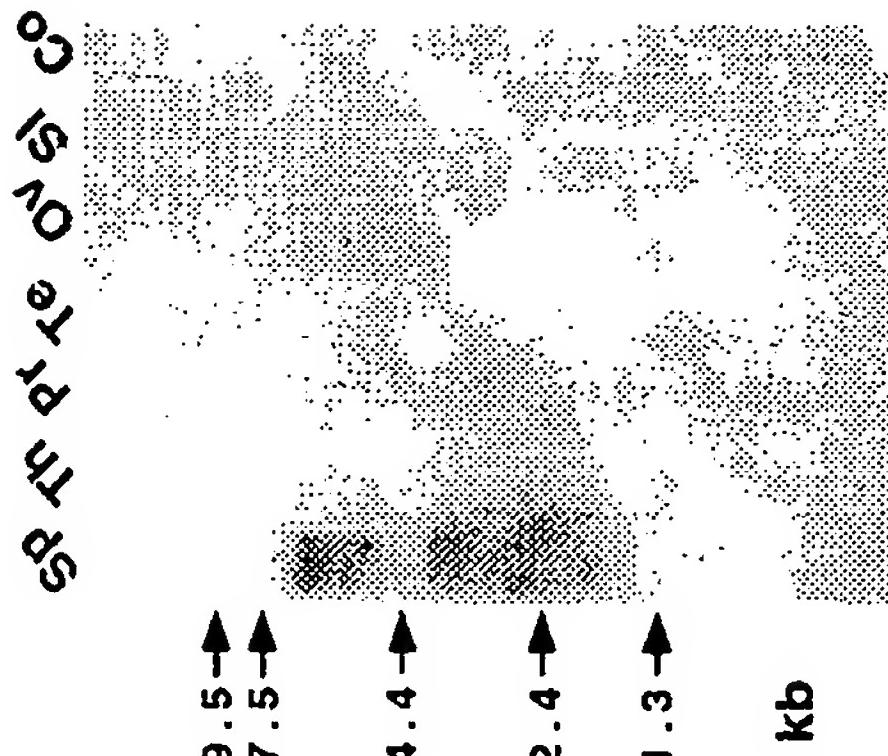


FIG. 5F